# TRANSACTIONS

OF THE

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

# AMERICAN MATHEMATICAL SOCIETY

EDITED BY

DUNHAM JACKSON

FRANCIS R. SHARPE

HOWARD HAWKS MITCHELL

WITH THE COÖPERATION OF

| | | |
|---|---|---|
| RAYMOND W. BRINK | EDWARD W. CHITTENDEN | JESSE DOUGLAS |
| WILLIAM C. GRAUSTEIN | OLIVE C. HAZLETT | EINAR HILLE |
| AUBREY J. KEMPNER | JOHN R. KLINE | CHARLES N. MOORE |
| GEORGE Y. RAINICH | JOSEPH F. RITT | CAROLINE E. SEELY |
| CHARLES H. SISAM | JACOB D. TAMARKIN | J. H. M. WEDDERBURN |

VOLUME 32

1930

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1930

General
Bound December 1930

COMPOSED, PRINTED AND BOUND BY
The Collegiate Press
GEORGE BANTA PUBLISHING COMPANY
MENASHA, WISCONSIN

43371

QA
1
F 001
v. 32

# TABLE OF CONTENTS

## VOLUME 32, 1930

PAGE

AGNEW, R. P., of Ithaca, N. Y. The behavior of bounds and oscillations of sequences of functions under regular transformations . . . . . 669

ALBERT, A. A., of New York, N. Y. New results in the theory of normal division algebras . . . . . 171

BAMFORTH, F. R., of Ithaca, N. Y. Surface transformations . . . . . 417

BAMFORTH, F. R., and BIRKHOFF, G. D., of Cambridge, Mass. Divergent series and singular points of ordinary differential equations . . . . 114

BELL, E. T., of Pasadena, Calif. Numbers of representations of integers in a certain triad of ternary quadratic forms . . . . . 1

——— Analytic functions in the irregular field of all numerical functions . 869

BIRKHOFF, G. D., and BAMFORTH, F. R., of Cambridge, Mass. Divergent series and singular points of ordinary differential equations . . . . 114

CALUGARÉANO, G., of Bucarest, Roumania. On differential equations admitting polygenic integrals . . . . . 110

CARLITZ, L., of Philadelphia, Pa. On Galois fields of certain types . . 451

COOK, A. J., of Edmonton, Canada. Pairs of rectilinear congruences with generators in one-to-one correspondence . . . . . 31

DICKSON, L. E., of Chicago, Ill. Construction of division algebras . . . 319

DYE, L. A., of Ithaca, N. Y. Involutorial transformations in $S_3$ of order $n$ with an $(n-1)$-fold line . . . . . 251

EARL, J. M., of Iowa City, Iowa. Polynomials of best approximation on aninfinite interval. . . . . 888

ENGSTROM, H. T., of New Haven, Conn. On the common index divisors of an algebraic field . . . . . 223

——— The theorem of Dedekind in the ideal theory of Zolotarev . . 879

FÉRAUD, L., of Cambridge, Mass. On Birkhoff's Pfaffian systems. . . 817

GHOSH, S., of Calcutta, India. On the solution of the equations of elastic equilibrium suitable for elliptic boundaries . . . . . 47

GILLESPIE, D. C., and HURWITZ, W. A., of Ithaca, N. Y. On sequences of continuous functions having continuous limits . . . . . 527

GOURIN, E., of New York, N. Y. On irreducible polynomials in several variables which become reducible when the variables are replaced by powers of themselves . . . . . 485

GROVE, V. G., of East Lansing, Mich. A canonical form of Green's projective analogue of the Gauss differential equations . . . . . 473

HAZLETT, O. C., of Urbana, Ill. On division algebras . . . . . 912

HOTELLING, H., of Stanford University, Calif. The consistency and ultimate distribution of optimum statistics . . . . . 847

HURWITZ, W. A., and GILLESPIE, D. C., of Ithaca, N. Y. On sequences of continuous functions having continuous limits. . . . . 527

LAMSON, K. W., of Bethlehem, Pa. Some differential and algebraic consequences of the Einstein field equations . . . . . 709

PAGE

LANE, E. P., of Chicago, Ill. Hypergeodesic mapping of a surface on a plane . . . . . 558

——— Integral surfaces of pairs of partial differential equations of the third order . . . . . 782

LANGER, R. E., of Madison, Wis. The boundary problem associated with a differential system rational in the parameter. . . . . 238

LA PAZ, L., of Chicago, Ill. An inverse problem of the calculus of variations for multiple integrals . . . . . 509

LATIMER, C. G., of Lexington, Ky. On certain identities in theta functions . . . . . 832

MAIER, W. E., of Chicago, Ill. Elementary properties of the $t_r$-functions . 905

MARDEN, M., of Cambridge, Mass. On the zeros of linear partial fractions . . . . . 81

——— On the zeros of certain rational functions . . . . . 658

MOORE, C. N., of Cincinnati, Ohio. On Gibbs's phenomenon for the developments in Bessel's functions . . . . . 409

MORSE, M., of Cambridge, Mass. The foundations of a theory of the calculus of variations in the large in $m$-space (second paper) . . . 599

POOR, V. C., of Ann Arbor, Mich. Residues of polygenic functions . . 216

POST, E. L., of New York, N. Y. Generalized differentiation . . . . 723

REID, W. T., of Austin, Texas. Properties of solutions of an infinite system of ordinary linear differential equations of the first order with auxiliary boundary conditions . . . . . 284

RITT, J. F., of New York, N. Y. On a certain ring of functions of two variables . . . . . 155

——— Manifolds of functions defined by systems of algebraic differential equations . . . . . 569

ROBERTS, J. H., of Austin, Texas. Concerning non-dense plane continua . 6

ROTH, W. E., of Milwaukee, Wis. On the unilateral equation in matrices . 61

TAMARKIN, J. D., of Providence, R. I. The notion of the Green's function in the theory of integro-differential equations . . . . . 860

TRJITZINSKY, W. J., of Bethlehem, Pa. On composition of singularities . 196

VANDIVER, H. S., of Austin, Texas. On power characters of singular integers in a properly irregular cyclotomic field . . . . . 391

WALSH, J. L., of Cambridge, Mass. Boundary values of an analytic function and the Tchebycheff method of approximation . . . . . 335

——— On the overconvergence of sequences of polynomials of best approximation. . . . . 794

WARD, L. E., of Iowa City, Iowa. Series associated with certain irregular third-order boundary value problems . . . . . 544

WARD, M., of Pasadena, Calif. Postulates for the inverse operations in a group . . . . . 520

WEISS, M. J., of Chicago, Ill. The limit of transitivity of a substitution group . . . . . 262

TABLE OF CONTENTS

PAGE

WHYBURN, G. T., of Austin, Texas. Cut points of connected sets and of continua . . . . . 147

——— On the structure of connected and connected im kleinen point sets . . . . . 926

WHYBURN, W. M., of Los Angeles, Calif. Functional properties of the solutions of differential systems . . . . . 502

WILDER, R. L., of Ann Arbor, Mich. A converse of the Jordan-Brouwer separation theorem in three dimensions . . . . . 632

ERRATA, volumes 26 and 31 . . . . . 944

# NUMBERS OF REPRESENTATIONS OF INTEGERS IN A CERTAIN TRIAD OF TERNARY QUADRATIC FORMS*

BY
E. T. BELL

## I. Introduction and summaries

1. A powerful method concerning numbers of representations of integers in positive definite quadratic or bilinear forms is illustrated by a complete discussion for the three forms

$$xy + yz + zx, \quad xy + yz + 2zx, \quad xy + 2yz + 2zx.$$

The results for these forms are intrinsically interesting. It is a feature of the method that, in general, at least two forms are handled simultaneously. Complete results are attainable for all of the forms in a given set or for none. We are interested here only in obtaining final results involving nothing more abstruse than the binary quadratic class number. For the possibilities sketched in §7 the class number is insufficient.

2. Without further remark, $a$, $b$, $c$ denote constant integers $>0$, and, unless otherwise stated, $x, y, z$ denote integers ranging $\geqq 0$; $n$ is an arbitrary (odd or even) constant integer $>0$; $m$ is an arbitrary odd constant integer $>0$; $\alpha$ is an arbitrary constant integer $\geqq 0$, so that $2^\alpha m$ is an arbitrary constant integer $>0$. If restrictions be imposed upon $m$, as for example $m \equiv 3 \bmod 8$, they will be indicated. Similarly for $x, y, z$.

The number of matrices $(x, y, z)$ such that $n = axy+byz+czx$ is called, as usual, the number of representations of $n$ in the form $axy+byz+czx$. This number will be denoted by $N_0[n=axy+byx+czx]$. The number of solutions $(x, y, z)$ for which $x>0$, $y>0$, $z>0$, will be denoted by $N_1[n = axy+byz+czx]$. Let $\zeta(x) \equiv 0$, or the number of divisors of $x$, according as $x$ is not, or is, an integer $>0$. Then, evidently,

$$\zeta(2^\alpha m) = (\alpha + 1)\zeta(m) ;$$

$$N_0[n = axy + byz + cx] = N_1[n=axy+byz+czx]+\zeta(n/a)+\zeta(n/b)+\zeta(n/c).$$

In particular

$$N_0[n = xy + yz + 2zx] = N_1[n = xy + yz + 2zx] + 2\zeta(n) + \zeta(n/2),$$
$$N_0[n = xy + 2yz + 2zx] = N_1[n = xy + 2yz + 2zx] + 2\zeta(n/2) + \zeta(n).$$

---

* Presented to the Society, November 29, 1929; received by the editors, September 14, 1929.



BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

3. The numbers $N_j (j=0, 1)$ have hitherto been determined only in the case $(a, b, c) = (1, 1, 1)$. As the results for this case will be required we summarize them.*

As always henceforth, $F(n)$ is the number of uneven classes of binary quadratic forms for the determinant $-n$, with all the conventions of H. J. S. Smith's *Report on the Theory of Numbers.*†

SUMMARY A

$$N_j(n) \equiv N_j[n = xy + yz + zx] \qquad (j = 0,1).$$

$$N_j(2^{2\alpha+1}m) = 3[(2^{\alpha+1} - 1)F(2m) + (-1)^j(\alpha + 1)\zeta(m)].$$

$$m \equiv 1 \bmod 4: \quad N_j(2^{2\alpha}m) = 3[(2^{\alpha+1} - 1)F(m) + (-1)^j\tfrac{1}{2}(2\alpha + 1)\zeta(m)].$$

$$m \equiv 3 \bmod 8: \quad N_j(2^{2\alpha}m) = 3[2(2^{\alpha} - \tfrac{1}{3})F(m) + (-1)^j\tfrac{1}{2}(2\alpha + 1)\zeta(m)].$$

$$m \equiv 7 \bmod 8: \quad N_j(2^{2\alpha}m) = 3[2^{\alpha+1}F(m) + (-1)^j\tfrac{1}{2}(2\alpha + 1)\zeta(m)].$$

These come at once from the formulas XIV quoted, or from Gage's form, on using the following easily seen reduction formulas, in which $G(n)$ is the whole number of classes of binary quadratic forms for the determinant $-n$.

$$G(2^{2\alpha+1}m) = (2^{\alpha+1} - 1)F(2m); \quad F(2^{2\alpha}m) = 2^{\alpha}F(m); \quad F(2^{2\alpha+1}m) = 2^{\alpha}F(2m).$$

$$m \equiv 1 \bmod 4: \quad G(2^{2\alpha}m) = (2^{\alpha+1} - 1)F(m).$$

$$m \equiv 3 \bmod 8: \quad G(2^{2\alpha}m) = 2(2^{\alpha} - \tfrac{1}{3})F(m).$$

$$m \equiv 7 \bmod 8: \quad G(2^{2\alpha}m) = 2^{\alpha+1}F(m).$$

As implied, the cases $(a, b, c) = (1, 1, 1), (1, 1, 2), (1, 2, 2)$ are intimately connected by the method which we shall indicate. For the last two we find the following.

SUMMARY B

$$N_j(n) \equiv N_j[n = xy + yz + 2zx] \qquad (j = 0,1).$$

$$N_j(m) = F(2m) + (-1)^j\zeta(m).$$

$$N_j(2^{2\alpha+2}m) = (2^{\alpha+3} - 3)F(2m) + (-1)^j(3\alpha + 4)\zeta(m).$$

* See Dickson, *History of the Theory of Numbers*, vol. 3, p. 224; Gage, American Journal of Mathematics, vol. 51 (1929), pp. 345–348. In the last, make the obvious correction on p. 348 of $\frac{1}{2}(\alpha+1)$ for $(\alpha+\frac{1}{2})$. Summary A contains the formulas which I suppressed in my paper, Tôhoku Mathematical Journal, vol. 19 (1921), pp. 105–116, as being (obviously) "rather less simple" than the elegant theorem given there for prime values of $n$. Until after seeing Mordell's paper, American Journal of Mathematics, vol. 45 (1923), pp. 1–4, I did not notice that all cases can be included in a single formula, say that given by Gage. The results follow even more simply than in the last from the formulas XIV of the paper cited in §6. There is now no need to reproduce this derivation, as it is superseded by those of Mordell and Gage.

† Mathematical Papers, vol. 1. We have used the elementary reduction formulas on p. 323. See also Dickson's *History*, vol. 3, chapter 6, where the final notation of Kronecker, used here, is explained.

$$m \equiv 1 \bmod 4: \quad N_j(2^{2\alpha+1}m) = (2^{\alpha+3} - 3)F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 5)\zeta(m).$$
$$m \equiv 3 \bmod 8: \quad N_j(2^{2\alpha+1}m) = 2(2^{\alpha+2} - 1)F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 5)\zeta(m).$$
$$m \equiv 7 \bmod 8: \quad N_j(2^{2\alpha+1}m) = 2^{\alpha+3}F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 5)\zeta(m).$$

## SUMMARY C

$$N_j(n) \equiv N_j[n = xy + 2yz + 2zx] \qquad (j = 0, 1).$$
$$N_j(m) = \tfrac{1}{2}[2F(m) + (-1)^j\zeta(m)].$$
$$N_j(2m) = 2[F(m) + (-1)^j\zeta(m)].$$
$$N_j(2^{2\alpha+3}m) = [2^{\alpha+2} + 3(2^{\alpha+1} - 1)]F(2m) + (-1)^j(3\alpha + 5)\zeta(m).$$
$$m \equiv 1 \bmod 4: \quad N_j(2^{2\alpha+2}m) = [2^{\alpha+2} + 3(2^{\alpha+1} - 1)]F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 7)\zeta(m).$$
$$m \equiv 3 \bmod 8: \quad N_j(2^{2\alpha+2}m) = 2[2^{\alpha+1} + 3\cdot 2^{\alpha} - 1]F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 7)\zeta(m).$$
$$m \equiv 7 \bmod 8: \quad N_j(2^{2\alpha+2}m) = 5\cdot 2^{\alpha+1}F(m) + (-1)^j\tfrac{1}{2}(6\alpha + 7)\zeta(m).$$

We do not express the foregoing in terms of $G$, as $F$ is the simplest of the class number functions and is that usually tabulated.

4. The results in Summaries B, C follow immediately from Summary A by means of the theorems next stated.

THEOREM 1. $N_1[2^{\alpha}m = xy + yz + 2zx,\ y\ odd] = F(2^{\alpha+1}m) - \zeta(m)$.

THEOREM 2. $N_1[m = xy + yz + 2zx] = F(2m) - \zeta(m)$.

This is the obvious corollary for $\alpha = 0$ of Theorem 1.

THEOREM 3. $N_1[m = xy + 2yz + 2zx] = \tfrac{1}{2}[2F(m) - \zeta(m)]$.

These, with many more which, however, are irrelevant for Summaries B, C, are given by the algebra in §6 by which these will be instantaneously proved.

## II. PROOFS OF SUMMARIES B, C

5. Assuming for a moment the results stated in Summary A and Theorems 1–3, all of which, as will be seen in §6, follow from the same simple analysis, we shall deduce Summaries B, C. As stated in §2, footnote, it is unnecessary to give details for Summary A.

Consider $N_1(n)$ in Summary C. In $N_1[2^{\alpha+2}m = xy + 2yz + 2zx]$, $xy$ must be even. Hence we may have $x$ even, $y$ odd, or $x$ odd, $y$ even, or $x$ even, $y$ even. Thus the $N_1$ in question is equal to

$$N_1[2^{\alpha+2}m = 2xy + 2yz + 4zx,\ y\ \text{odd}] + N_1[2^{\alpha+2}m = 2xy + 4yz + 2zx,\ x\ \text{odd}] + N_1[2^{\alpha+2}m = 4xy + 4yz + 4zx],$$

and therefore is equal to

$$2N_1[2^{\alpha+1}m = xy + yz + 2zx, y \text{ odd}] + N_1[2^\alpha m = xy + yz + zx].$$

Similarly,

$$N_1[2m = xy + 2yz + 2zx] = 2N_1[m = xy + yz + 2zx],$$

since on the left either but not both of $x$, $y$ may be even; also

$$N_1[2^{\alpha+1}m = xy + xz + 2zx] = N_1[2^\alpha m = xy + yz + zx] + N_1[2^{\alpha+1}m = xy + yz + 2zx, y \text{ odd}].$$

The rest follows from the above by obvious substitutions and reductions in Summary A and Theorems 1–3.

## III. Proofs of Theorems 1–3

6. Let $f(x)$ denote a function which is uniform and finite for all integer values $\geqq 0$ of $x$, and which, for such values of $x$, satisfies $f(x)=f(-x)$. Consider all the solutions $(t_1, \tau_1, d_2, \delta_2)$, in integers $>0$, where $t_1$, $\tau_1$ are odd, $d_2\,\delta_2$ is even, and $d_2$, $\delta_2$ are of opposite parities, of $m=t_1\tau_1+d_2\delta_2$. Then

$$2\sum F(m-a^2)f(2a) = \zeta(m)f(0) + 2\sum\left[\sum_{r=1}^{(r-1)/2} f(2r)\right] + \sum'(\delta-d)f(\delta+d) + \sum'[f(\delta_2+d_2-\tau_1) - f(\delta_2+d_2+\tau_1)],$$

in which $\sum$ on the left refers to $a=0, \pm1, \pm2, \cdots, \pm[m^{1/2}]$, where $[x]$ is the greatest integer in $x$; $\sum$ on the right refers to all divisors $\tau>0$ of $m$; the first $\sum'$ refers to all divisors $d$, $\delta>0$ of $m$ such that $m=d\delta$, $d<\delta$, and the second $\sum'$ refers to all sets $(\delta_2, d_2, \tau_1)$.

In this formula take for $f(x)$ the special case $f(0)=1$, $f(x)=0$, $x\neq0$. A straightforward interpretation of the result gives Theorem 3. In precisely the same way Theorem 1 follows from the next.

Consider all the solutions, for $l$ constant, $>0$, and even, of $l=l_2+l_3$, where $l_2$, $l_3$ are $>0$ and even, and (all letters denote integers $>0$), $l=t\tau=d\delta=d_1\delta_1$; $d$, $\delta$ are of like parities, $d<\delta$; $\tau$ is odd; $d_1$, $\delta_1$ are of opposite parities; $l_2=t_2\tau_2$, $\tau_2$ is odd; $l_3=d_3\delta_3$, and $d_3$, $\delta_3$ are of opposite parities:

$$l = t_2\tau_2 + d_3\delta_3.$$

Then

$$4\sum F(l-a^2)f(2a) = 2\sum\left[f(0) + 2\sum_{r=0}^{(r-1)/2} f(2r)\right] + \sum'\left[f(0) + 2\left[\sum_{r=1}^{(\delta_1+d_1-1)/2} f(2r)\right]\right] + 2\sum'(\delta-d)f(\delta+d) + 2\sum'[f(\delta_3+d_3-\tau_2) - f(\delta_3+d_3+\tau_2)],$$

the $\sum$ on the left referring to $a=0, \pm 1, \pm 2, \cdots, \pm [l^{1/2}]$.

The stated $f$-theorems are numbers $XII_1$, $XII_2$ of a former paper on singly infinite class number relations.* As already stated, the formulas XIV of that paper give immediately (inter alia) Summary A.

7. The method mentioned in the introduction is now sufficiently obvious. It may be of interest to indicate a few equally evident consequences of this procedure. First, it is clear that a change of $q$ into $-q$ in an elliptic theta identity from which a paraphrase is inferred, can give no new paraphrases; the arithmetical information contained in both is necessarily the same. Second, to introduce coefficients divisible by a given integer $b$ into the quadratic or bilinear forms relating to a given paraphrase, it is necessary and sufficient to apply to the original theta identity the transformation of order $kb$ ($k \equiv$ an integer) before paraphrasing or, what is equivalent, to replace $f(x)$ by $\cos xz$ after paraphrasing, and assign to $z$ successively the values $2\pi/\beta$, where $\beta$ is divisible only by those primes that divide $b$. Third, if instead of $f(x)$, the paraphrase concerns odd functions $g(x) = -g(-x)$, the resulting information, obtained on taking $g(x) = 1$, 0, or $-1$ according as $x>0$, $x=0$ or $x<0$, refers to differences of specialized numbers of representations in the same forms.

The sets of forms simultaneously treated are given by the transformation of any preassigned order.

It is clear that other choices of $f(x)$, as for example $f(x)=1$ (this does not apply in the present case, but it does in all paraphrases obtained from products of even theta quotients exclusively), or $f(x)=x^2$, or $x^4, \cdots$, give further theorems on numbers of representations. That all are merely aspects of single formulas illustrates the power of the method.

The number of theta expansions in the literature which give simple, interesting arithmetical results is not great. This number has recently been very considerably extended. The works cited put a large number of interesting arithmetical forms within range of exact treatment.†

* Quarterly Journal of Mathematics, vol. 49 (1923), pp. 322–337, especially p. 335.

† Three theses: D. A. F. Robinson (Chicago), for the 48 of the 64 functions $\theta_\alpha(x+y)/[\theta_\beta(x)\theta_\gamma(x)]$ not already expanded by Jacobi, Hermite, Krause and others; J. D. Elder, and M. A. Basoco (California Institute of Technology), who have greatly amplified the work of Biehler, Hermite, Appell, Krause and others on doubly periodic functions of the third kind. Abstracts of the last two are given in the Proceedings of the National Academy of Sciences, vol. 15 (1929), No. 8, pp. 668–671, 675–677.

California Institute of Technology,
Pasadena, Calif.

# CONCERNING NON-DENSE PLANE CONTINUA*

BY

J. H. ROBERTS

It has been shown by Menger† that a necessary and sufficient condition that a plane continuum $M$ contains no domain is that for each point $P$ of $M$ and each positive number $\epsilon$ there exists a simple closed curve $J$ of diameter less than $\epsilon$ which encloses $P$, and such that $M \cdot J$ is totally disconnected.

In the present paper it is shown that if $M$ is a continuum which contains no domain then there exists a set $G$ of simple closed curves filling the whole plane and indeed topologically equivalent to the set of all polygons, such that the common part of $M$ and any curve of the set $G$ is vacuous or totally disconnected. Additional results are obtained for the special case where $M$ is a continuous curve.

I wish to acknowledge my indebtedness to Professor R. L. Moore, and to thank him. Credit is due him for the suggestion of most of the theorems of this paper, and for many helpful criticisms of the proofs.

## PART I

In this paper I make frequent use of the notion of a *double ruling*.‡ If, on the simple closed curve $ABCDA$, $X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m, X_n', X_{n-1}', \cdots, X_1', Y_m', Y_{m-1}', \cdots, Y_1'$ are points in the order $AX_1X_2 \cdots X_nBY_1Y_2 \cdots Y_mCX_n'X_{n-1}' \cdots X_1'DY_m'Y_{m-1}' \cdots Y_1'A$, and $X_1X_1', X_2X_2', \cdots, X_nX_n'$, and $Y_1Y_1', Y_2Y_2', \cdots, Y_mY_m'$ are arcs which, except for their end points, lie entirely within $ABCDA$, and finally, for every $i, j (1 \leqq i \leqq n, 1 \leqq j \leqq m)$ $X_iX_i'$ has just one point in common with $Y_jY_j'$ and no point in common with $X_jX_j'$ (unless $i=j$), then these two sets of arcs are said to constitute a *double ruling* of the interior of $ABCDA$ (or merely a *double ruling* of $ABCDA$). The arcs $X_1X_1', X_2X_2', \cdots, X_nX_n'$ are said to be *parallel* to $BC$ and to $AD$, and the arcs $Y_1Y_1', Y_2Y_2', \cdots, Y_mY_m'$ are said

* Presented to the Society, in part, December 28, 1927, and December 27, 1928; received by the editors March 30, 1929.

† *Über die Dimensionalität von Punktmengen*, Monatshefte für Mathematik und Physik, vol. 33 (1923), pp. 148–160. In his paper *Sur les multiplicités Cantoriennes*, Fundamenta Mathematicae, vol. 7 (1925), pp. 30–137, Urysohn obtained the slightly weaker result that for each point $P$ of $M$ and each positive number $\epsilon$ there exists a totally disconnected closed subset $T$ of $M$ such that $M-T$ is the sum of two mutually separated sets $M_1$ and $M_2$, such that $M_1$ contains $P$ and is of diameter less than $\epsilon$.

‡ See R. L. Moore, *Concerning a set of postulates for plane analysis situs*, these Transactions, vol. 20 (1919), p. 172.

to be *parallel* to $AB$ and $CD$. Each of these two sets of arcs is a *single ruling* of $ABCDA$. Two sets of arcs $H$ and $K$ are said to constitute a *complete double ruling* of $ABCDA$ if (1) every finite subset of the set of arcs $H+K$ is a single or double ruling of $ABCDA$,(2) through each point within $ABCDA$ there is an arc of $H$ and an arc of $K$, and (3) each point of $ABCDA$ (except $A$, $B$, $C$ and $D$) is an end point of some arc of $H$ or of some arc of $K$. If $P$ is a point with both coördinates rational, then $P$ is said to be a rational point. By a rational line is meant a line whose equation is $x=r$ or $y=r$ where $r$ is some rational number.

As an obvious corollary of a theorem due to Fréchet* I state the following

THEOREM I. *If $M$ is a point set containing no domain then there exists a continuous transformation $T$ of the plane $S$ into itself such that if $P$ is any point of $M$ then $T(P)$ is not a rational point.*

Schoenflies has proved the following theorem:†

*If $T_1$ is a continuous one-to-one correspondence between the points of two simple closed curves $J_1$ and $J_2$ such that $T_1(J_1)=J_2$, then there exists a continuous one-to-one correspondence $T_2$ between $J_1$ plus its interior and $J_2$ plus its interior such that $T_2(P)=T_1(P)$ for every point $P$ on $J_1$.*

Therefore for every simple closed curve $J$ there exists a continuous transformation $T$ which throws $J$ plus its interior into a square plus its interior. The truth of the following lemma is apparent in view of this fact and Theorem I.

LEMMA I. *Suppose $M$ is a point set containing no rational point, and $A$, $B$, $C$ and $D$ are rational points which are the vertices of a rectangle, and $AB$ is a horizontal interval. Suppose $\alpha$ is a double ruling of $ABCD$ with arcs $h_1, h_2, \cdots, h_n$ parallel to $BA$ and arcs $v_1, v_2, \cdots, v_m$ parallel to $BC$ such that* (1) *for each $i$ [$j$] ($i \leqq n, j \leqq m$) the end points of $h_i[v_j]$ have the same rational $y$-coördinate [$x$-coördinate],* (2) *for every $i$ and $j$ ($i \leqq n, j \leqq m$) the point common to $h_i$ and $v_j$ is rational, and* (3) *the rational points are everywhere dense on $h_i$ and on $v_j$ ($i \leqq n$, $j \leqq m$). Then there exists a continuous transformation $T$ throwing $ABCD$ plus its interior into itself and such that* (1) *$T$ reduces to the identity transformation on the rectangle $ABCD$,* (2) *$T(P)$ is not rational for any point $P$ of $M$, and* (3) *for every $i$[$j$] ($i \leqq n, j \leqq m$) the arc $T(h_i)$ [$T(v_j)$] is an interval of some horizontal [vertical] rational line.*

* Mathematische Annalen, vol. 68 (1910), p. 159. See also Urysohn, loc. cit., p. 83.

† *Beiträge zur Theorie der Punktmengen*, Mathematische Annalen, vol. 62 (1906), pp. 286–328. See also J. R. Kline, *A new proof of a theorem due to Schoenflies*, Proceedings of the National Academy of Sciences, vol. 6 (1920), pp. 529–531.

THEOREM II. *If $M$ is the sum of a countable number of closed point sets containing no domain and lying in a euclidean plane $S$, then there exists a continuous transformation $T$ of $S$ into itself such that if $L$ denotes any straight line of $S$ the point set $L \cdot T(M)$ is either totally disconnected or vacuous, and no point of $T(M)$ is rational.*

I will assume that $M$ contains no rational point. In view of Theorem I this is no essential restriction. Let $A, B, C, D$ and $E$ be points with coördinates

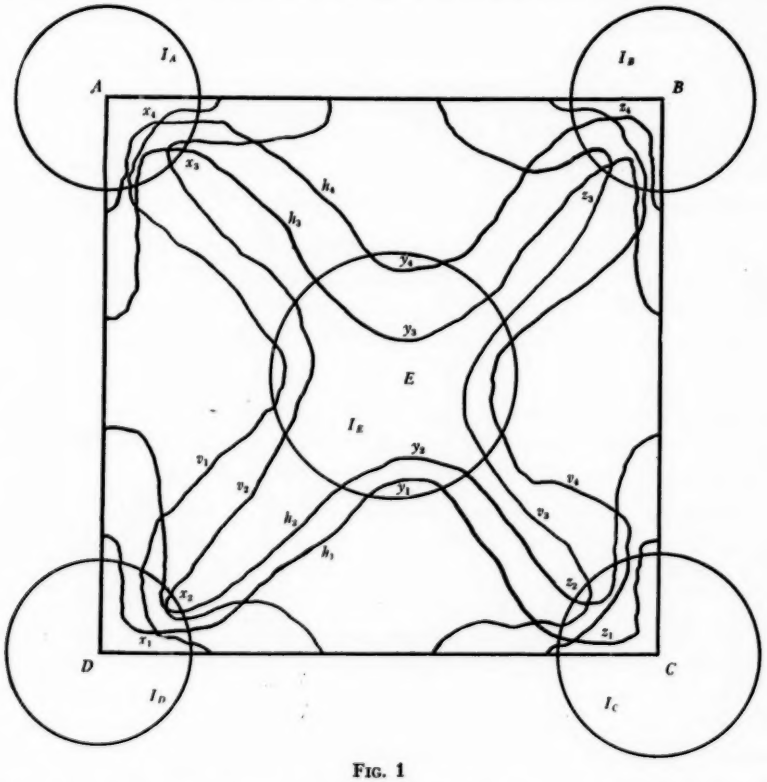


FIG. 1

(0,1), (1,1), (1,0), (0,0) and (1/2, 1/2) respectively, and let $R$ denote the interior of the square $ABCD$. Suppose $M = M_1 + M_2 + M_3 + \cdots$, where for every $n$ the set $M_n$ is closed. There exist five circles with centers $A, B, C, D$ and $E$ respectively, such that their interiors $I_A, I_B, I_C, I_D$ and $I_E$ are mutually exclusive, and no one of them contains a point of the closed set $M_1$.

There exists a double ruling $\alpha_1$ (see Fig. 1) of $R$ consisting of four arcs

$h_1, \cdots, h_4$ parallel to $AB$ and four arcs $v_1, \cdots, v_4$ parallel to $BC$ such that (1) the end points of the arcs $h_1, \cdots, h_4, v_1, \cdots, v_4$ divide each side of the square $ABCD$ into five *equal* parts, (2) for each $i(i \leqq 4)$ the arc $h_i$ contains three points, $x_i$, $y_i$ and $z_i$, in the order $x_iy_iz_i$ from $AD$ to $BC$ which belong to the domains $I_D$, $I_E$, and $I_C$ respectively, for $i=1, 2$, and to the domains $I_A$, $I_E$, and $I_B$ respectively for $i=3, 4$, (3) the points $x_1, x_2, x_3, x_4, z_1, z_2, z_3, z_4$ are the points $h_1 \cdot v_1$, $h_2 \cdot v_2$, $h_3 \cdot v_2$, $h_4 \cdot v_1$, $h_1 \cdot v_4$, $h_2 \cdot v_3$, $h_3 \cdot v_3$, $h_4 \cdot v_4$, respectively, (4) for each $i$ $(i \leqq 4)$ $v_i$ contains a point of $I_E$ between the arcs $h_2$ and $h_3$, and (5) for each $i$ the rational points are dense on $h_i$ and on $v_i$ and every point $h_i \cdot v_j$ $(i \leqq 4, j \leqq 4)$ is rational. Since $M$ does not contain any rational

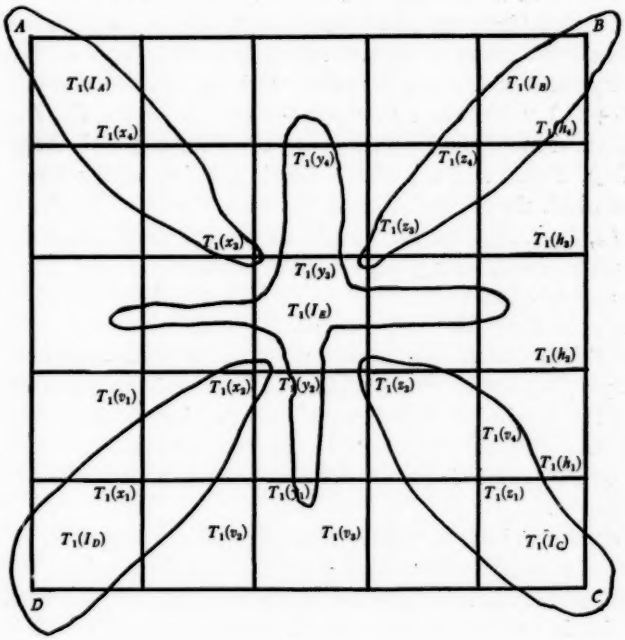


FIG. 2

point it is easily seen that the hypotheses of Lemma I are satisfied by $M$ and the double ruling $\alpha_1$. Let $T_1$ denote a transformation satisfying the conclusion of Lemma I. It is easily seen that every straight line which contains two points of the square $ABCD$ contains a point not belonging to $T_1(M_1)$. (See Fig. 2.)

Let $t_1$ be the set of all arcs $k$ such that $k=T_1(g)$, where $g$ is an arc of the

double ruling $\alpha_1$. Obviously $t_1$ is a double ruling of $R$, the arcs of which are horizontal and vertical rational intervals. There exists a positive number $\epsilon_1(\epsilon_1<1/2)$ such that if $P$ and $Q$ are points of $\overline{R}$ and $\delta(P, Q)$† $\geqq 1/2$, then $\delta[T_1(P), T_1(Q)]>\epsilon_1$. Let $P_1, P_2, P_3, \cdots$ denote the rational points on the arcs $AB$ and $BC$. Let $\beta_1$ denote a double ruling of $R$ consisting of a finite set of rational horizontal and vertical intervals such that (1) $\beta_1$ contains an interval through the point $P_1$, and contains‡ $t_1$ and (2) every component§ of $\overline{R}$ minus the sum of the arcs of the ruling $\beta_1$ is of diameter less than $\epsilon_1$.

Let $F_1$ denote the collection of all components of $R$ minus the sum of the arcs of the ruling $\beta_1$ and let $G$ denote any element of $F_1$. The domain $G$ is the interior of a rectangle $A_GB_GC_GD_G$, where $A_GB_G$ is an interval of a horizontal rational line, and all of the points $A_G$, $B_G$, $C_G$, and $D_G$ are rational. Let $E_G$ denote the center of the rectangle and let $I_{A_G}$, $I_{B_G}$, $I_{C_G}$, $I_{D_G}$, and $I_{E_G}$ be mutually exclusive circular domains which contain no point of the closed set $T_1(M_1+M_2)$ and whose centers are the points $A_G$, $B_G$, $C_G$, $D_G$, and $E_G$. Clearly there exists a double ruling $\alpha_{2_G}$ of $G$ having with respect to $\overline{G}$, $T_1(M_1+M_2)$, $T_1(M)$ and the above mentioned circular domains the same properties which the double ruling $\alpha_1$ has with respect to $\overline{R}$, $M_1$, $M$ and the domains $I_A$, $I_B$, $I_C$, $I_D$ and $I_E$. Let $N_1$ be the sum of all arcs $k$ where $k$ belongs to some double ruling $\alpha_{2_G}$ for some domain $G$ of $F_1$. The point set $N_1$ is the sum of a finite number of arcs which constitute a double ruling $\gamma_1$ of $R$. Let $\alpha_2$ be the double ruling consisting of all arcs of the rulings $\beta_1$ and $\gamma_1$. Since $T_1(M)$ contains no rational point the hypotheses of Lemma I are satisfied by $T_1(M)$ and the double ruling $\alpha_2$. There exists a transformation $T_2$ satisfying the conclusion of Lemma I and reducing to the identity transformation on the square $ABCD$ and on the arcs of the ruling $\beta_1$. Let $t_2$ denote the double ruling consisting of all arcs $k$ such that $k=T_2(g)$, where $g$ is an arc of the ruling $\alpha_2$.

Let $r$ denote an interval of length $2\epsilon_1$ which is a subset of $R$. Clearly if $r$ is a subset of some arc of the ruling $\beta_1$ then it contains a point not belonging to $T_2T_1(M_1+M_2)$. If $r$ is not a subarc of any arc of $\beta_1$ then there exists a rectangle which is a subset of the sum of the arcs of $\beta_1$ and which contains exactly two points of $r$. With the assistance of Fig. 2 it can readily be seen that $r$ contains a point not belonging to the set $T_2T_1(M_1+M_2)$. Hence if $T$ is any continuous transformation of $\overline{R}$ into itself such that $T(P)=P$ for every point $P$ on $ABCD$ or on some arc of the ruling $t_2$, then every interval in $\overline{R}$

† If $P$ and $Q$ are points, $\delta(P, Q)$ will denote the distance from $P$ to $Q$.

‡ A double ruling $\beta$ is said to *contain* a double ruling $t$ if every arc of $t$ is an arc of $\beta$.

§ By a *component* of a point set $G$ is meant a maximal connected subset of $G$.

which is of length greater than or equal to $2\epsilon_1$ contains a point not belonging to the set $TT_2T_1(M_1+M_2)$. There exists a positive number $\epsilon_2$ ($\epsilon_2<1/3$) such that if $\delta(U, V)\geqq 1/3$ and $U$ and $V$ belong to $\overline{R}$, then $\delta[T_2T_1(U), T_2T_1(V)]>\epsilon_2$. Let $\beta_2$ denote a double ruling of $R$ consisting of rational horizontal and vertical intervals such that (1) some interval of $\beta_2$ contains $P_2$ (the second rational point on $ABCD$), (2) $\beta_2$ contains $t_2$, (3) every component of $R$ minus the sum of the arcs of the ruling $\beta_2$ is of diameter less than $\epsilon_2$.

Let $F_2$ denote the collection of all components of $R$ minus the sum of the arcs of the ruling $\beta_2$ and let $G$ denote any element of $F_2$. The domain $G$ is the interior of a rectangle composed of subsets of rational lines. There exists a double ruling $\alpha_{3_G}$ of $G$ having with respect to $\overline{G}$, $T_2T_1(M_1+M_2+M_3)$ and $T_2T_1(M)$ the properties which the double ruling $\alpha_1$ has with respect to $\overline{R}$, $M_1$ and $M$. Let $N_2$ be the sum of all arcs $k$, where $k$ belongs to some double ruling $\alpha_{3_G}$ for some domain $G$ of $F_2$. The point set $N_2$ is the sum of a finite number of arcs which constitute a double ruling $\gamma_2$ of $R$. Let $\alpha_3$ be the double ruling consisting of all arcs of the rulings $\beta_2$ and $\gamma_2$. Since $T_2T_1(M)$ contains no rational point the hypotheses of Lemma I are satisfied by $T_2T_1(M)$ and the double ruling $\alpha_3$. Hence there exists a transformation $T_3$ satisfying the conclusion of Lemma I and reducing to the identity transformation on $ABCD$ and on the arcs of the double ruling $\beta_2$. Let $t_3$ denote the double ruling consisting of all arcs $k$ such that $k=T_3(g)$, where $g$ is an arc of the ruling $\alpha_3$. Each component of $\overline{R}$ minus the sum of the arcs of the ruling $\beta_2$ is of diameter less than $\epsilon_2$. Hence it follows that if $T$ is a continuous transformation of $\overline{R}$ into itself which reduces to the identity transformation on $ABCD$ and on every arc of the ruling $t_3$ then every interval which is a subset of $\overline{R}$ and is of diameter greater than or equal to $2\epsilon_2$ contains a point not belonging to the set $TT_3T_2T_1(M_1+M_2+M_3)$.

Proceeding in this way one can see that there exists a countable infinity of double rulings $t_1, t_2, \cdots$, and a countable infinity of continuous transformations $T_1, T_2, T_3, \cdots$ of $\overline{R}$ into itself such that for every positive integer $n$ the following properties obtain: (1) $t_{n+1}$ contains $t_n$, (2) $t_n$ is composed of horizontal and vertical *rational* intervals and every component of $\overline{R}$ minus the sum of the arcs of the ruling $t_{n+1}$ is of diameter less than $1/n$, (3) if $U^m$ denotes the point $T_mT_{m-1}\cdots T_1(U)$ and $U$ and $V$ are points of $\overline{R}$ such that $\delta(U, V)\geqq 1/n$, then for every integer $m(m>n)$ the points $U^m$ and $V^m$ are separated in $\overline{R}$ by some arc of the ruling $t_n$, and, for every $n$, $\delta(U^{n+1}, U^{n+k})<2/n$ ($k=1, 2, \cdots$), (4) the transformation $T_{n+1}$ reduces to the identity transformation on $ABCD$ and on all arcs of the ruling $t_n$, (5) some arc of $t_n$ contains $P_n$ (the $n$th rational point on $ABCD$), (6) no point belonging to two arcs of $t_n$ belongs to the set $T_nT_{n-1}\cdots T_1(M)$, and (7) if $T$ is any con-

tinuous transformation of $\overline{R}$ into itself which reduces to the identity transformation on $ABCD$ and on the arcs of the ruling $t_{n+1}$ then every interval which is a subset of $\overline{R}$ and is of length greater than or equal to $2/n$ contains a point not belonging to the set $TT_{n+1}T_n \cdots T_1(M_1+M_2+\cdots+M_{n+1})$.

From property (3) it clearly follows that for every point $U$ of $\overline{R}$ the sequence $U, U^1, U^2, \cdots$ has a sequential limit point. Let $T$ be the transformation which carries $U$ into this sequential limit point. In particular, if for some integer $n$ the point $U^n$ belongs to some arc of $t_n$ or to $ABCD$, then $U^n = U^{n+k}$ $(k=1, 2, \cdots)$ and $T(U)$ is merely $U^n$. I shall now show that $T$ is a continuous one-to-one transformation of $\overline{R}$ into itself which satisfies the conclusion of Theorem II with respect to $\overline{R}$.

First, $T$ is a one-to-one transformation of $\overline{R}$ into itself which reduces to the identity transformation on $ABCD$. Let $U$ and $V$ denote distinct points of $\overline{R}$. We see at once from property (3) that there exist two mutually exclusive intervals $g_1$ and $g_2$ which belong to some double ruling $t_n$ and each of which separates $U^m$ from $V^m$ in $\overline{R}$ for every integer $m(m>n)$. Then the sequences $U^1, U^2, U^3, \cdots$ and $V^1, V^2, V^3, \cdots$ have distinct sequential limit points. Hence for each point $X$ of $\overline{R}$ there is not more than one point $U_X$ such that $X=T(U_X)$. That for each $X$ there is at least one point $U_X$ follows from properties (2) and (3).

Second, the transformation $T$ is continuous. Let $U$ be a point not belonging to a point set $W$. If $U$ is not a limit point of $W$ then there exists an integer $m$ such that any point of $T_nT_{n-1}\cdots T_1(W)$ is separated from $T(U)$ by some arc of the ruling $t_n$ for every $n$ $(n>m)$. Hence $T(U)$ is not a limit point of $T(W)$. If however $U$ is a limit point of the set $W$ then for every $\epsilon$ there is a point $V_\epsilon$ of $W$ and an integer $n_\epsilon$ such that $\delta(U^m, V_\epsilon^m)<\epsilon$ for every integer $m$ greater than $n_\epsilon$. Hence $T(U)$ is a limit point of $T(W)$.

It follows from properties (2), (5) and (6) that if $U$ is a point such that $T(U)$ is rational, then $U$ does not belong to $M$.

Let $L$ denote any interval which is a subset of $\overline{R}$, and suppose that $L\cdot T(M)$ contains a nondegenerate† connected set. Since no continuum is the sum of a countable number of totally disconnected closed sets it follows that there exists an integer $n$ such that $L$ contains a nondegenerate connected subset of $T(M_n)$. Let $r$ denote some interval which is a subset of $L\cdot T(M_n)$ and let $\epsilon$ be its length. Let $n_1$ be an integer such that $2/n_1<\epsilon$ and $n_1>n$. Now the transformation $T$ reduces to the transformation $T_{n_1+1}T_{n_1}\cdots T_1$ on $ABCD$ and on every arc of $t_{n_1}$ (property 4). Hence if $T^*$

† A point set containing but a single point is said to be degenerate.

denotes the transformation such that $T(U)=T^*T_{n_1+1}T_{n_1}\cdots T_1(U)$ then $T^*$ reduces to the identity transformation on $ABCD$ and on every arc of $t_{n_1}$. Hence by property (7) the interval $r$ contains a point not belonging to the set $T^*T_{n_1+1}T_{n_1}\cdots T_1(M_1+\cdots+M_{n_1+1})$. But this set is exactly $T(M_1+\cdots+M_{n_1+1})$. Hence we have a contradiction which means that no interval which is a subset of $\overline{R}$ contains a nondegenerate connected subset of $T(M)$.

I have now shown that $T$ is a continuous one-to-one transformation of $\overline{R}$ into itself such that if $L$ is any straight line then the point set $L\cdot T(M)\cdot\overline{R}$ is either vacuous or totally disconnected, and no point which beongs to $T(M)$ is rational. Hence $T$ satisfies the conclusion of Theorem II with respect to the rectangle $ABCD$ plus its interior. The extension to the whole plane is obvious.

## PART II

In his paper† *Grundzüge einer Theorie der Kurven*, Karl Menger proves that if $M$ is a bounded continuum containing no domain then a necessary and sufficient‡ condition that $M$ be a continuous curve [regular curve§] is that, for each positive number $\epsilon$, $M$ is the sum of a finite number of continua all of diameter less than $\epsilon$, such that the common part of any two of these continua is vacuous or totally disconnected [finite]. The necessity of this condition follows from Theorem III for the case where $M$ is a continuous curve, and from Theorem VI for the case where $M$ is a regular curve.

THEOREM III. *If $M$ is any continuous curve which contains no domain then there exists a continuous transformation $T$ of the plane $S$ into itself such that* (1) *no straight line contains a nondegenerate connected subset of $T(M)$ and no point of $T(M)$ is rational, and* (2) *if $R$ is the interior of a rectangle composed of intervals of lines with equations of the form $x=r$ and $y=r$, where $r$ is a rational number, then the point set $R\cdot T(M)$ is the sum of a finite number of connected sets and every point of $M$ on the boundary of $R$ is a limit point of $R\cdot T(M)$.*

To help establish Theorem III I will first prove the following lemma.

---

† Mathematische Annalen, vol. 95 (1925), pp. 277–306.

‡ For the case of a continuous curve the sufficiency of the given condition was proved by W. Sierpinski, *Sur une condition pour qu'un continu soit une courbe jordanienne*, Fundamenta Mathematicae, vol. 1 (1920), pp. 44–60.

§ A continuum $M$ is said to be a *regular curve* if for each point $P$ of $M$ and each positive number $\epsilon$ there exists a domain of diameter less than $\epsilon$ which contains $P$ and whose boundary contains only a finite number of points of $M$. See K. Menger, loc. cit.

LEMMA II. *If $M$ is a continuous curve containing no domain and $J$ is a simple closed curve $EKFGLHE$ such that the arcs $EKF$ and $GLH$ of $J$ contain no point of $M$, then there exists a simple continuous arc $KL$ lying, except for $K$ and $L$, within $J$, and such that* (1) *the arc $KL$ contains no nondegenerate connected subset of $M$,* (2) *every point of $M$ on the arc $KL$ is a limit point of $M$ from both sides† of $KL$, and* (3) *the point set $M-M\cdot KL$ is the sum of a finite number of connected sets.*

**Proof of Lemma II.** Let $I$ denote the interior of $J$ and let $N$ be the point set $M\cdot I+EH+FG$. If $N$ is not connected there exists an arc $KL$ which contains no point of $M$. I will therefore suppose that $N$ is connected. Then $N$ is a continuous curve. The set of junction‡ points of $M$ is countable.‡ With the help of this fact and Theorem I it is easily seen that there exists an arc $KZL$ which except for $K$ and $L$ is within $J$ and which contains no junction point of $M$ and no nondegenerate connected subset of $M$. Let $S_E$ and $\overline{S}_F$ denote the components of $N-N\cdot KZL$ containing $E$ and $F$ respectively. Let $S_E^*$ denote the component of $N-\overline{S}_F$ which contains $E$. Let $S_F^*$ denote $N-\overline{S}_E^*$.

Since $N$ is a continuous curve, $S_E^*$ and $S_F^*$ are mutually separated sets. The common part of the two sets $\overline{S}_E^*$ and $\overline{S}_F^*$ is a subset of $N\cdot KL$ and is therefore totally disconnected. Now $S_E^*$ is connected, by definition. I will show that $S_F^*$ is also connected.

Suppose that $Q$ is a point of $S_F$ and $P$ is any point of $S_F^*-S_F^*\cdot\overline{S}_F$. There exists§ in $N$ a simple continuous arc $PQ$. Let $P_1$ be the first point on this arc from $P$ to $Q$ which belongs to the set $\overline{S}_E^*+\overline{S}_F$. The point $P_1$ belongs to $\overline{S}_F$, for otherwise the subarc $PP_1$ of $PQ$ belongs to $S_E^*$, which is contrary to the supposition that $P$ is in $S_F^*$. If $P_1$ is not in $\overline{S}_E^*$ then it is in $\overline{S}_F$ and the arc $PP_1$ is connected to $Q$ by an arc in $S_F^*$.

Suppose then that $P_1$ belongs to both of the sets $\overline{S}_E^*$ and $\overline{S}_F$. Then $P_1$ is a

---

† If $M$ is a point set and $P$ is an interior point of an arc $AB$ then $P$ is said to be a limit point of $M$ from both sides of $AB$ if there exists a simple closed curve $J$ containing $AB$ such that, $I_J$ denoting the interior of $J$, $P$ is a limit point of $M\cdot I_J$ and of $M\cdot(S-\bar{I}_J)$.

‡ Cf. R. L. Moore, *Concerning triods in the plane and the junction points of plane continua*, Proceedings of the National Academy of Sciences, vol. 14 (1928). If $P$ is a point of a continuous curve $N$, and $K$ is a domain containing $P$ such that $P$ is a cut point of the component of $N\cdot K$ which contains $P$, and furthermore there exist three arcs $PA_1$, $PA_2$, and $PA_3$ which lie in $N$ and have only the point $P$ in common, then $P$ is said to be a *junction point* of $N$. The continuum $PA_1+PA_2+PA_3$ is called a *triod* and the point $P$ is its *emanation point*.

§ R. L. Moore, *A theorem concerning continuous curves*, Bulletin of the American Mathematical Society, vol. 23 (1917), pp. 233–236. See also Mazurkiewicz, *Sur les lignes de Jordan*, Fundamenta Mathematicae, vol. 1 (1920), pp. 166–209, and H. Tietze, *Ueber stetige Kurven, Jordansche Kurvenbogen, und geschlossene Jordansche Kurven*, Mathematische Zeitschrift, vol. 5 (1919), pp. 284–291.

limit point of each of the mutually separated connected sets $S_E^*$ and $S_F$. Therefore† there exists an arc $R_1P_1$ in $S_E^*+P_1$, and an arc $T_1P_1$ in $S_F+P_1$, where $R_1$ and $T_1$ are on the arcs $EH$ and $FG$, respectively. (See Fig. 3.) Since the point $P_1$ is an emanation point of the triod of $N$ composed of the three arcs $R_1P_1$, $PP_1$, and $T_1P_1$, and is not a junction point of $M$, it is not a

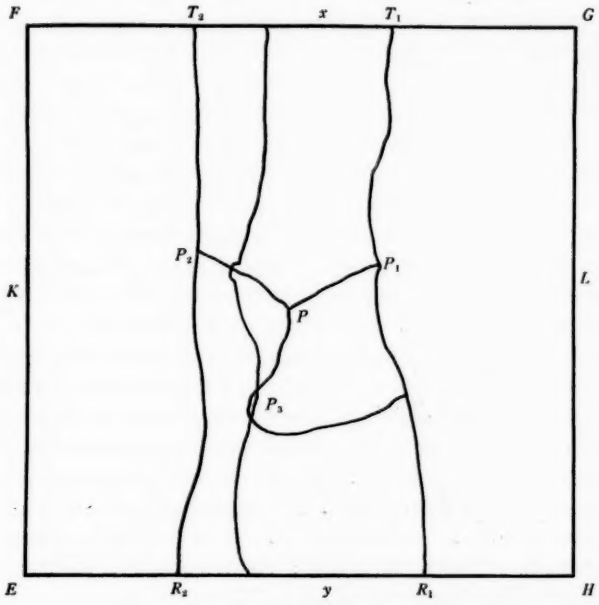


FIG. 3

cut point of $N$. Therefore there exists an arc from $P$ to $Q$ which does not contain the point $P_1$. Let $P_2$ denote the first point of this arc in the order from $P$ to $Q$ which belongs to the set $\overline{S}_E^*+\overline{S}_F$. Then as before either $P$ and $Q$ can be connected by an arc in $S_F^*$ or $P_2$ belongs to both $\overline{S}_E^*$ and $\overline{S}_F$. Suppose the latter is true. Then $P_2$ does not belong to any of the arcs $PP_1$, $R_1P_1$ or $T_1P_1$. In $S_F+P_2$ there exists an arc $P_2T_2$, where $T_2$ is either on the arc $FG$ or on the arc $P_1T_1$ and no other point of the arc $P_2T_2$ belongs to either $FG$ or $P_1T_1$. In $S_E^*+P_2$ there exists an arc $P_2R_2$, where $R_2$ is on $EH$ or on $R_1P_1$, and no other point of $P_2R_2$ is on $EH$ or $R_1P_1$. Suppose for definiteness that $T_2$ and $R_2$ are on the arcs $FG$ and $EH$, respectively. Let $X$ and $Y$ be points

† See R. L. Wilder, *Concerning continuous curves*, Fundamenta Mathematicae, vol. 7 (1925), pp. 340–377.

on $FG$ and $EH$ between $T_1$ and $T_2$, and $R_1$ and $R_2$, respectively. Then the simple closed curve $h$ ($h=T_1XT_2P_2R_2YR_1P_1T_1$) encloses the segments $PP_1$ and $PP_2$. Let $R$ be a domain which contains the arc $PP_1$ but does not contain the point $P_2$. The component of $N\cdot R$ which contains $P_1$ contains subarcs of $PP_1$, $R_1P_1$, and $T_1P_1$. Since $P_1$ is not a junction point of $N$ it is not a cut point of this component. Hence there exists within $R\cdot(N-P_1)$ a simple continuous arc $PW$, where $W$ belongs to one of the segments $R_1P_1$ and $T_1P_1$. Let $P_3$ be the first point in the order from $P$ to $W$ which the arc $PW$ has in common with the set $\overline{S}_E^*+\overline{S}_F$. Assume that $P_3$ belongs to both $\overline{S}_E^*$ and $\overline{S}_F$. Clearly $P_3$ is within the simple closed curve $h$. There exists a segment of an arc within $h$ with $P_1$ and $P_2$ as end points which contains no point of $\overline{S}_E^*+\overline{S}_F$. This segment $P_1P_2$ divides the interior of $h$ into two connected domains, one of which contains $P_3$. Now there exist arcs from $P_3$ to $T_1$ and from $P_3$ to $R_1$ which lie in the sets $S_F+P_3$ and $S_E^*+P_3$, respectively. Since the simple closed curve $P_1R_1YR_2P_2P_1$ contains no point of $S_F+P_3$, and the simple closed curve $P_1T_1XT_2P_2P_1$ contains no point of $S_E^*+P_3$ it is clear that we have reached a contradiction. Hence every point $P$ of $S_F^*-S_F^*\cdot\overline{S}_F$ lies in the component of $S_F^*$ which contains $S_F$, which means that $S_F^*$ is connected.

We now have $N=\overline{S}_E^*+\overline{S}_F^*$ where $S_E^*$ and $S_F^*$ are connected and have no point in common, and $\overline{S}_E^*\cdot\overline{S}_F^*$ is totally disconnected. Let $X_E$ and $X_F$ be the two continua obtained by adding to $\overline{S}_E^*$ and $\overline{S}_F^*$, respectively, all of their bounded complementary domains. Now no point is in a bounded complementary domain of both $\overline{S}_E^*$ and $\overline{S}_F^*$. Therefore the point set $X_E\cdot X_F$ is the same as the set $\overline{S}_E^*\cdot\overline{S}_F^*$. Call this set $T$. Then $X_E-T$ is connected, and neither $X_E$ nor $X_F$ separates the plane. As a result of a theorem of R. L. Moore† it follows that there exists a simple closed curve $k$ enclosing $X_E-T$, containing $T$, and not containing or enclosing any point of $X_F-T$. Clearly $k$ contains an arc whose end points lie on the segments $EF$ and $GH$, respectively, but which otherwise lies within $J$. This arc can be modified so as to have $K$ and $L$ for end points and retain the property of separating $X_E-T$ and $X_F-T$ as above.‡ I will show that this arc satisfies the conclusion of Lemma II.

(1) Clearly the arc $KL$ contains no nondegenerate connected subset of $M$, for it contains no point of $M$ not belonging to the totally disconnected set $T$. (2) Every point of $M$ on the arc $KL$ is a limit point of each of the sets $S_E^*$ and $S_F^*$ and is therefore a limit point of $M$ from both sides of $KL$.

† *Concerning the separation of point sets by curves*, Proceedings of the National Academy of Sciences, vol. 11 (1925), pp. 469–476.

‡ This follows readily from the fact that the segments $EKF$ and $GLH$ do not contain any point of $N$.

(3) Every component of $M-M\cdot KL$ contains a point either on the arc $EH$ or on the arc $FG$ because of the fact that $(M-T)+EH+FG$ is the sum of two or less connected sets. Since not infinitely many components of $M-T$ have a point on $EH$ or $FG$ and also have a limit point on $KL$ it follows that only a finite number of such components can have a limit point on $KL$. But *every* component of $M-T$ has a limit point on $KL$. Hence the number of such components is finite. This completes the proof of Lemma II.

**Proof of Theorom III.** As in Theorem II there is no loss of generality in assuming that no point with both coördinates rational belongs to $M$. Clearly the plane $S$ can be regarded as the sum of a countable infinity of rectangles plus their interiors, the rectangles being subsets of *irrational* horizontal and vertical lines. Let $R_1, R_2, R_3, \cdots$ denote the mutually exclusive interiors of such a set of rectangles.

If $\epsilon$ is any positive number and if we have any finite double ruling $\alpha$ of $R_i$ ($i$ being any positive integer) such that no arc of $\alpha$ contains a nondegenerate connected subset of $M$, then we can obtain a finite double ruling $\beta$ which contains the arcs of $\alpha$ and is such that every component of $R_i$ minus the sum of the arcs of $\beta$ is of diameter less than $\epsilon$, and every arc of $\beta$ which does not belong to $\alpha$ has properties (1), (2), and (3) of Lemma II, and such that no point common to two such arcs belongs to $M$. Now the arcs of the double ruling $\beta_n$ defined in the proof of Theorem II were taken to be *rational* lines so that no nondegenerate connected subset of $T_nT_{n-1}\cdots T_1(M)$ would be a subset of an arc of $\beta_n$. In view of Theorem II, however, it can be seen that transformations $T_1, T_2, T_3, \cdots$ can now be chosen so that no nondegenerate connected subset of $T_nT_{n-1}\cdots T_1(M)$ is a subset of *any* straight line. Hence some of the arcs of the rulings $\beta_n$ can be taken as *irrational* horizontal and vertical intervals. By obvious modification of the argument given in the proof of Theorem II it can be seen that there exists a countable infinity of double rulings $t_1, t_2, t_3, \cdots$ of $R_i$, and a countable infinity of continuous transformations $T_1, T_2, T_3, \cdots$ of $\overline{R}_i$ into itself which, except for (6) and a modification of (2) to allow $t_n$ to contain intervals of *irrational* horizontal and vertical lines, have properties (1)-(7) as stated in the proof of Theorem II, and the additional property that the double ruling $t_n$ contains a double ruling $r_n$ every arc of which has properties (1), (2), and (3) of Lemma II and $P_n$ (the $n$th rational point on the boundary of $R_i$) is an end point of some arc of the ruling $r_n$. Let $W_i$ be the transformation corresponding to the transformation $T$ as defined in the proof of Theorem II. Let $T$ be the transformation of the plane $S$ into itself which for every $i$ reduces to $W_i$ over $\overline{R}_i$.

Clearly then no straight line contains a nondegenerate connected sub-

set of $T(M)$, and no point of $T(M)$ is rational. The second conclusion of Theorem III follows readily from the fact that every rational horizontal and vertical interval with rational end points has properties (1), (2), and (3) of Lemma II with respect to the continuous curve $T(M)$. Hence Theorem III is established.

The following is an example of a regular curve which contains an uncountable set $H$ of points such that no arc containing a point of $H$ has property (3) of Lemma II. It therefore follows that Theorem III would be false if the stipulation that the boundary of $R$ is composed of intervals of *rational* horizontal and vertical lines were omitted, or if the word *rational* were replaced by the word *irrational*.

EXAMPLE 1. (See Fig. 4.) Let $H$ denote a nondense perfect point set on the interval $0 \leqq x \leqq 1$, and let $K$ denote any acyclic† continuous curve such that $H$ is the set of end points† of $K$. Let $G_1, G_2, G_3, \cdots$ denote a

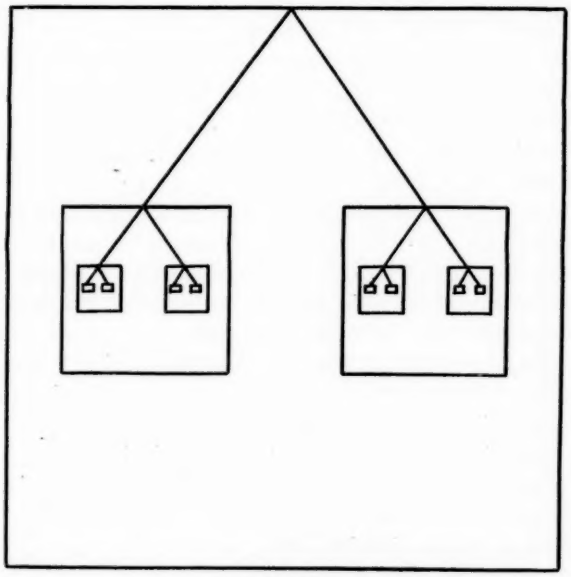

FIG. 4

† A continuous curve is said to be *acyclic* if it contains no simple closed curve. See H. M. Gehman, *Concerning acyclic continuous curves*, these Transactions, vol. 29 (1927), pp. 553–568. An end point of an acyclic continuous curve is a point which is not an interior point of any arc of that curve. See R. L. Wilder, loc. cit., or H. M. Gehman, loc. cit.

contracting sequence† of mutually exclusive simple closed curves such that every point of $H$ is enclosed by infinitely many curves of the sequence $G_1, G_2, G_3, \cdots$, and for every $n$ the curve $G_n$ has just one point in common with $K$, and this point does not belong to $H$. Let $M$ be the continuous curve $K+G_1+G_2+G_3+\cdots$. Let $P$ denote a point of $K-H$. Then if $Q$ is any point of $H$ there is one and only one arc $PQ$ from $P$ to $Q$ in $M$. Let $J$ denote a simple closed curve containing $Q$ but not containing a non-degenerate connected subset of $M$. Either (1) $PQ-PQ\cdot J$ is the sum of infinitely many maximal connected subsets, in which case $M-M\cdot J$ is not the sum of a finite number of connected sets, or (2) there exists a point $X$ on the arc $PQ$ distinct from $Q$ such that the arc $XQ$ has only the point $Q$ on $J$. Suppose for definiteness that $X$ is within $J$. There exist infinitely many simple closed curves of $M$ enclosing $Q$, having points within $J$ and points without $J$, and having only one point on the arc $XQ$ of $PQ$. Since no maximal connected subset of $M-M\cdot J$ which lies in the exterior of $J$ can contain a point of the arc $XQ$ it follows that the number of components of $M-M\cdot J$ is infinite.

In his paper‡ *Concerning irreducible cuttings of continua*, G. T. Whyburn raises the question as to whether or not every open§ subset of a plane continuous curve $M$ contains an irreducible cutting of $M$. This question is answered by the following theorem which is an application of Theorem III.

THEOREM IV. *Every open subset of a plane continuous curve $M$ contains an irreducible cutting of $M$.*

Let $G$ denote an open subset of a continuous curve $M$. Clearly if $G$ contains a domain then it contains a circle which is an irreducible cutting of $M$. If $G$ contains no domain let $R$ denote the interior of a circle such that $R$ contains a point of $G$ but $\overline{R}$ does not contain a point of $M-G$. Let $M_1$ denote any maximal connected subset of $M\cdot\overline{R}$ which contains more than one point. Then‖ $M_1$ is a continuous curve which contains no domain. From Theorem III it readily follows that there exists a simple closed curve $J$ which encloses some point of $M_1$ but does not contain or enclose any point

† If $H$ is a sequence of point sets and for each positive number $\epsilon$ only a finite number of point sets of the set $H$ are of diameter greater than $\epsilon$ then $H$ is said to be a *contracting sequence* of point sets. See R. L. Moore, *Concerning upper semi-continuous collections*, Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 81–88.

‡ Fundamenta Mathematicae, vol. 13, pp. 42–57.

§ An open subset of a continuum $M$ is a set such that its complement with respect to $M$ is closed. An irreducible cutting of a continuum $M$ is a point set $K$ of $M$ such that $M-K$ is not connected, but such that if $G$ is any proper subset of $K$ then $M-G$ is connected. See G. T. Whyburn, ibid.

‖ H. M. Gehman, *Concerning the subsets of a plane continuous curve*, Annals of Mathematics, vol. 27, p. 34.

of $M-M_1$, and such that the point set $M \cdot J$ is totally disconnected and separates $M_1$ into a finite number (greater than 1) of connected sets. Clearly then $M \cdot J$ has the same properties with respect to $M$. Then† the set $M \cdot J$ contains a subset which is an irreducible cutting of $M$.

THEOREM V. *A necessary and sufficient condition that a continuum $M$ (not the whole plane) be a regular curve is that if $R$ is a connected domain containing two distinct points $A$ and $B$ not belonging to $M$ then in $R$ there exists a simple continuous arc from $A$ to $B$ which contains only a finite number of points of $M$.*

The condition is necessary. Suppose $M$ is a regular curve and $R$ is a connected domain containing two points $A$ and $B$ not belonging to $M$. Let $AB$ denote any simple continuous arc from $A$ to $B$ which lies in $R$, and let $A'$ and $B'$ be points in the order $AA'B'B$ such that no point of $M$ is on the arc $AA'$ or the arc $BB'$ of $AB$. Enclosing each point of the arc $A'B'$ there exists a simple closed curve containing only a finite number of points of $M$ and not containing or enclosing $A$ or $B$ or any point not in the domain $R$. There exists a finite set of such curves whose interiors cover the arc $A'B'$. Call the curves of such a set $J_1, J_2, \cdots, J_n$. If $H$ denotes the continuous curve $AA'+BB'+J_1+J_2+\cdots+J_n$ then $H$ contains only a finite number of points of $M$. Let $AXB$ denote an arc from $A$ to $B$ which is a subset of $H$. Obviously this arc contains only a finite number of points of $M$.

The condition is sufficient. Clearly $M$ cannot contain a domain. Suppose $P$ is any point of $M$ and $\epsilon$ is any positive number. Let $J_1$ and $J_2$ denote two circles with $P$ as center and radii $\epsilon/2$ and $\epsilon/3$, respectively. Let $P_1$ and $P_2$ denote the extremities of a diameter of $J_1$. Let $D$ denote the domain bounded by $J_1+J_2$, and let $A$ and $B$ denote two points not belonging to $M$ and lying in $D$ on different sides of the diameter $P_1PP_2$. Let $D_1$ and $D_2$ be the connected domains $D-D \cdot PP_1$ and $D-D \cdot PP_2$, respectively. Let $AX_1B$ and $AX_2B$ denote arcs lying in $D_1$ and $D_2$, respectively, and containing only a finite number of points of $M$. The continuous curve $AX_1B+AX_2B$ contains a simple closed curve which encloses $P$, contains only a finite number of points of $M$, and is of diameter less than $\epsilon$. Hence the point $P$ is a regular point and $M$ is a regular curve.

THEOREM VI. *If $M$ is a regular curve (not necessarily bounded) in a euclidean plane $S$, then there exists a continuous transformation $T$ of $S$ into itself such that* (1) *no straight line contains a nondegenerate connected subset of $T(M)$ and no point of $T(M)$ is rational, and* (2) *each rational horizontal or vertical line has in common with $T(M)$ a point set which has no limit point.*

† See G. T. Whyburn, loc. cit.

The only essential difference between the proof of this theorem and that of Theorem III is that here we require that the arcs of the double ruling $r_n$ shall have only a finite number of points in common with $M$, instead of requiring that they have properties (1), (2) and (3) of Lemma II. In case $M$ is a *bounded* regular curve the second conclusion of Theorem VI is equivalent to the statement that no rational line contains more than a finite number of points of $M$.

It follows that if $R$ is the interior of a rectangle whose sides are intervals of rational horizontal and vertical lines, then the point set $T(M)\cdot(\overline{R}-R)$ contains only a finite number of points of $M$. However the following example shows that it does not follow that the set $R\cdot T(M)$ is the sum of a finite number of connected sets.

EXAMPLE 2. (See Fig. 5.) For each pair of positive integers $n$ and $k(k\leqq 2^n)$ let $I_{kn}$ denote the interval with end points $[(k-1)/2^n, 0]$ and $[k/2^n, 0]$, and let $C_{kn}$ denote the semicircle above the $x$-axis with $I_{kn}$ as diameter. Let $H$ denote the continuum which is the sum of the interval $I(0\leqq x\leqq 1)$ and all semicircles $I_{kn}(k\leqq 2^n,\ n=1,\ 2,\ 3,\cdots)$. Let $P_1,\ P_2,\ P_3,\cdots$ denote the points of the $x$-axis which are extremities of diameters of semicircles belonging

FIG. 5

to $H$, and for each $n$ let $a_{1n},\ a_{2n},\ a_{3n},\cdots$ denote a contracting sequence of arcs all of diameter less than $1/n$, such that for each $m$ the arc $a_{mn}$ contains the point $P_n$ but no other point of $H$ and no other point of the arc $a_{kn}$ $(k\neq m)$. Let $M$ be the continuum $H+\sum_{i=1}^{\infty}\sum_{n=1}^{\infty}a_{in}$. Then $M$ is a regular curve. Now any arc which lies between the lines $x=0$ and $x=1$ and has a point above and a point below the $x$-axis either contains the point $P_n$ for some integer $n$ or it contains infinitely many points of $M$. In the first case it cuts $M$ into infinitely many components.

THEOREM VII. *If every point of a bounded regular curve $M$ in a euclidean plane $S$ is of finite order† then there exists a continuous transformation $T$ of $S$ into itself such that* (1) *no straight line contains a nondegenerate connected subset*

† If $P$ is a point of a regular curve $M$ and there exists an integer $n$ such that for every positive number $\epsilon$ there is a domain of diameter less than $\epsilon$ which contains $P$ and whose boundary has not more than $n$ points in common with $M$ then the point $P$ is said to be of *finite order*. Cf. K. Menger, loc. cit.

*of $T(M)$ and no point of $T(M)$ is rational, (2) the rational lines contain only a finite number of points of $T(M)$, and (3) if $R$ is the interior of a rectangle composed of intervals of rational lines, then $R \cdot T(M)$ is the sum of a finite number of connected sets.*

Let $T$ be a transformation satisfying the conclusion of Theorem VI. Since each point of $M$ on a rational line $L$ is of finite order it is not a limit point of infinitely many components of $T(M)-L$. In view of this, and the additional fact that the set of points of $M$ on any rational line is finite, it is clear that the transformation $T$ satisfies the conclusion of Theorem VII.

THEOREM VIII. *If $M$ is a bounded regular curve which contains only a finite number of simple closed curves then there exists a transformation $T$ satisfying the conclusion of Theorem* VII.

To help prove Theorem VIII I will establish the following lemma.

LEMMA III. *If $M$ is a bounded regular curve which contains only a finite number of simple closed curves, and $R$ is a connected domain containing two points $A$ and $B$ not belonging to $M$, then there exists a simple continuous arc $AB$ which lies within $R$, contains only a finite number of points of $M$, and is such that $M-M \cdot AB$ is the sum of a finite number of connected sets.*

**Proof of Lemma III.** Let $H$ denote the set of junction points of $M$. Let $J$ be a simple closed curve enclosing $A$ and $B$ and lying in $R$, and let $g$ denote a simple continuous arc from $A$ to $B$ which lies within $J$, contains no non-degenerate connected subset of $M$, and no point of $H$. But the outer boundary of every bounded complementary domain of a continuous curve is† a simple closed curve and‡ no two bounded complementary domains of a continuous curve have the same outer boundary. Hence since $M$ contains only a finite number of simple closed curves it follows that only a finite number of complementary domains of $M+J$ have boundary points on the arc $g$. If one of these domains contains both $A$ and $B$ the lemma is obviously established. If not let $P_1$ be the last point of $g$ in the order from $A$ to $B$ which is on the boundary of that complementary domain of $M+J$ which contains $A$. Then§ there exists an arc $AP_1$ which lies wholly in this domain except for the point $P_1$. Since $P_1$ is a limit point of the points of $S-M$ on the

† R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), Theorem 4 and p. 259.

‡ R. L. Moore, *Concerning paths that do not separate a given continuous curve*, Proceedings of the National Academy of Sciences, vol. 12 (1926), Theorem 1.

§ Schoenflies, *Die Entwickelung der Lehre von den Punktmannigfaltigkeiten*, zweiter Teil, Jahresbericht der Deutschen Mathematiker-Vereinigung, Erganzungsbände, vol. 2 (1908).

arc $P_1B$ of $g$ it follows that $P_1$ is a boundary point of some complementary domain of $M+J$ which contains points on the arc $P_1B$. Let $P_2$ be the last point on the arc $P_1B$ belonging to the boundary of a complementary domain of $M+J$ which also has $P_1$ on its boundary. If $P_2$ is the same as $P_1$ then the arc $P_1B$ contains no point of $M+J$. In either case there exists a simple continuous arc with $P_1$ and $P_2$ as end points which contains no point of $M+J$ except $P_1$ and $P_2$. Continuing this process a finite number of times one obtains a simple continuous arc $AB$ which contains only a finite number of points of $M$ and no junction point of $M$. Clearly this arc satisfies the conclusion of the lemma.

A proof of Theorem VIII can now be given which is closely analogous to the proof of Theorem III. The essential difference is that the arcs of the double ruling $r_n$ are here to be chosen so as to have the properties stated in Lemma III rather than those stated in Lemma II.

THEOREM IX. *If $A$ and $B$ are distinct points of a continuous curve $M$ then $M$ contains a simple continuous arc from $A$ to $B$ every subarc of which contains a subarc which either lies on the boundary of some complementary domain of $M$ or lies in some domain which belongs to $M$.*

(1) Suppose $M$ contains no domain. Let $AXB$ denote a simple continuous arc from $A$ to $B$ such that the common part of $AXB$ and $M$ is totally disconnected. Let $T$ denote the set $M \cdot AXB$. Let $D_1, D_2, \cdots$ denote the complementary domains of $M$ which contain limit points on the arc $AXB$ and for each $i$ let $J_i$ denote the boundary of $D_i$. Let $K$ be the point set $T+J_1+J_2+J_3+\cdots$. Since $J_1, J_2, J_3, \cdots$ is a contracting sequence of continuous curves all containing points on the arc $AXB$ it is readily seen that the set $K$ is closed. If $P$ is an interior point of the arc $AXB$ which does not belong to $K$ then there exists a connected subset of $K$ containing the last point of $T$ which precedes $P$ on the arc $AXB$ and the first point of $T$ which follows $P$ on this arc. Therefore $K$ is connected. With the use of the fact that the boundary of every complementary domain of a continuous curve is itself a continuous curve† it readily follows that $K$ is connected im kleinen. Hence $K$ is a continuous curve. Let $AB$ denote any arc which lies in $K$‡ and let $EF$ denote any subarc of $AB$. Since $T$ is totally disconnected the arc $EF$ contains a subarc $E'F'$ which contains no point of $T$. The arc $E'F'$ is a subset of $J_1+J_2+J_3+\cdots$, and hence is equal to $J_1 \cdot E'F'+J_2 \cdot E'F'+J_3 \cdot E'F'+\cdots$. But the sum of a countable number of totally dis-

† R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), p. 259.

‡ See third footnote on p. 14.

connected closed point sets is not connected. Hence there exists at least one integer $i$ such that the set $J_i \cdot E'F'$ contains an arc. Thus every subarc of $AB$ contains an arc belonging to the boundary of some complementary domain of $M$.

(2) If $A$ and $B$ are distinct points of a continuous curve $M$ which contains a domain let $M_1$ be a continuous curve containing $A$ and $B$ which is obtained by taking from $M$ the interiors $I_1, I_2, I_3, \cdots$ of a contracting sequence of circles such that (1) for every domain $D$ which is a subset of $M$ there is an integer $n$ such that $I_n$ contains at least one point of $D$, (2) for every $n$, $I_n$ is a subset of some domain belonging to $M$ and (3) $\bar{I}_k \bar{I}_n = 0 (k \neq n)$. Let $AB$ denote an arc satisfying the conclusion of the theorem with respect to $M_1$. Since the boundary of a complementary domain of $M_1$ which is not a complementary domain of $M$ belongs in a domain lying in $M$ it is obvious that $AB$ satisfies the conclusion of the theorem with respect to $M$.

THEOREM X. *If $M$ is a bounded continuous curve which contains no domain then there exists a continuous transformation $T$ of the plane $S$ into itself such that* (1) *if $AB$ is an arc such that $T(AB)$ is a subset of a rational line then $AB \cdot M = c_1 + c_2 + \cdots + c_n$ where for each $i$ $(i \leqq n)$ $c_i$ is an arc or a point and if $c_i$ is an arc then every subarc of $c_i$ contains a subarc lying on the boundary of some complementary domain of $M$ and* (2) *if $AB$ is an arc such that $T(AB)$ is a subset of an irrational horizontal or vertical line then $AB \cdot M$ is vacuous or totally disconnected.*

To help establish Theorem X, I will prove several lemmas. To avoid repetition I will say that an arc $AB$ has *property $c_\epsilon$ with respect to $M$*, or merely that it has *property $c_\epsilon$* if the common part of $M$ and $AB$ is the sum of a finite number of connected sets such that each of these sets which is an arc is of diameter less than $\epsilon$ and has the property that every subarc of it contains a subarc lying on the boundary of some complementary domain of $M$.

LEMMA IV. *If $I$ is the interior of a simple closed curve and $M$ is a bounded continuous curve containing no domain and $A$ and $B$ are distinct points lying in $I$ and $\epsilon$ is any positive number, then there exists a simple continuous arc from $A$ to $B$ which is a subset of $I$ and which has property $c_\epsilon$.*

With the help of Theorem II it can readily be seen that there exists a simple closed curve $J_1$ lying in $I$, enclosing $A$ and $B$, and such that (1) $J_1 \cdot M$ is totally disconnected, and (2) if $I_1$ denotes the interior of $J_1$ then no component of $M \cdot \bar{I}_1$ is of diameter greater than or equal to $\epsilon$. Let $AXB$ denote any simple continuous arc from $A$ to $B$ which lies in $I_1$ and let $s_1, s_2, \cdots, s_n$ denote the components of $M \cdot \bar{I}_1$ which have points on $AXB$. For each $i$

$(i \leqq n)$ $s_i$ is† a continuous curve. In view of this fact and Theorem IX it follows that there exists an arc $AB$ in $I_1$ such that $M \cdot AB$ is a subset of $s_1+s_2+\cdots+s_n$ and is the sum of $n$ or less connected sets such that each of these sets which is an arc has the properties of the arc of Theorem IX with respect to that one of the continuous curves $s_1, s_2, \cdots, s_n$ to which it belongs. Let $EF$ denote an arc belonging to $M \cdot AB$. Since $EF$ contains a subarc lying wholly within $J$ it can easily be shown that $EF$ has the properties stated in Theorem IX with respect to $M$. Since in addition $EF$ is of diameter less than $\epsilon$ the lemma is proved.

LEMMA V. *If $J$ is a simple closed curve and $KL$ is a simple continuous arc which lies within $J$ except that $K$ and $L$ are on $J$ and $KL$ is on the boundary of a complementary domain $D$ of $M$, then there exists a simple continuous arc $AB$ which lies within $J$ such that* (1) *the common part of $AB$ and $KL$ is a single point,* (2) *$KL$ separates $A$ from $B$ within $J$, and* (3) *$AB \cdot M$ is either an arc or a point, and if it is an arc it has property $c_\epsilon$.*

Let $C$ denote the interior of a circle which lies within $J$ and encloses a point of $KL$. There exists a point $A$ in $C \cdot D$ and a subarc $E'F'$ of $EF$ such that for every point $P$ of $E'F'$ there exists an arc $AP$ which lies in $C \cdot D$ except for the point $P$. Let $O$ denote some interior point of $E'F'$ and let $C_1$ denote the interior of a circle $J_1$ which lies in $C$ such that $C_1$ contains $O$ but contains no point of $KL-E'F'$. Let $A'$ and $B'$ denote points in $C_1$ lying respectively on the $A$ side and the non $A$ side of $KL$ and let $B'A'$ denote an arc having property $c_\epsilon$ and lying in $C_1$. Let $Q$ denote the first point of $B'A'$ on $KL$ in the order from $B'$ to $A'$. Let $AQ$ denote an arc lying in $C \cdot D$ except for the point $Q$. Let $QB$ denote a subarc of $QB'$ such that $QB \cdot M$ is connected. The sum of the arcs $AQ$ and $QB$ gives an arc $AB$ which satisfies the conclusion of the lemma.

LEMMA VI. *If $M$ is a continuous curve containing no domain and lying within a simple closed curve $J$ whose interior is $R$, $\epsilon_1$ and $\epsilon_2$ are any positive numbers and $\alpha$ is a double ruling of $R$ such that every arc of $\alpha$ has property $c_{\epsilon_1}$, then there exists a double ruling $\beta$ of $R$ such that every arc of $\alpha$ is also an arc of $\beta$, every arc of $\beta$ which is not an arc of $\alpha$ has property $c_{\epsilon_2}$ and every component of $\overline{R}$ minus the sum of the arcs of the ruling $\beta$ is of diameter less than $\epsilon_2$.*

With the help of a theorem of Schoenflies‡ it is easily seen that there exists a continuous transformation $T_1$ of the plane into itself which throws $J$ into a square $ABCD$ and the arcs of $\alpha$ into horizontal and vertical inter-

† See H. M. Gehman, loc. cit.

‡ Loc. cit.

vals. Let $\alpha_1$ be the double ruling of $T(R)$ which is composed of all arcs $T(g)$ where $g$ is an arc of $\alpha$. Since $T_1$ is continuous it obviously follows that there exist two finite sets of rectangles $h_1, h_2, \cdots, h_n, v_1, v_2, \cdots, v_n$ such that (1) for each $i$ $(i \leqq n)$, $h_i[v_i]$ has no point in common with any arc of $\alpha_1$ which is parallel to $AB$ $[BC]$ and no point in common with $h_j[v_j]$ $(j \leqq n, j \neq i)$ but contains at least one point of $AD$ and one point of $BC$ $[AB$ and $CD]$, and (2) if $\alpha_2$ is any double ruling of $R$ which contains all of the arcs of $\alpha_1$ and in addition contains arcs $a_i$ and $b_i$ such that $T(a_i)$ lies in $h_i$ plus its interior and $T(b_i)$ lies in $v_i$ plus its interior $(i=1, 2, \cdots, n)$ then every component of $\overline{R}$ minus the sum of the arcs of $\alpha_2$ is of diameter less than $\epsilon_2$. It is easily shown with the help of Lemmas IV and V that a particular such ruling $\beta$ can be obtained such that the arcs of $\beta$ which do not belong to $\alpha$ have property $c_{\epsilon_2}$.

LEMMA VII. *Suppose $M$ is a continuous curve which contains no domain and lies in the interior $R$ of a square $ABCD$. Let $\alpha$ be any double ruling of $R$ and let $P$ be a point of $R$ not belonging to any arc of $\alpha$. Then there exists an integer $k$ such that if $n>k$ and $\beta$ is any double ruling of $R$ such that* (1) *$\beta$ contains $\alpha$,* (2) *no arc of $\beta$ contains $P$,* (3) *every component of $\overline{R}$ minus the arcs of $\beta$ is of diameter less than $1/n$, and* (4) *every arc of $\beta$ which does not belong to $\alpha$ has property $c_{1/n}$, then if $E$ denotes the component, containing $P$, of $\overline{R}$ minus the arcs of $\beta$ which are parallel to $AB$ $[BC]$, there exists in $E$ an arc $a_E$ with property $c_{1/(n+1)}$ which together with the arcs of $\beta$ forms a double ruling of $R$ and such that no component of $M \cdot (E - a_E)$ contains points in more than two components of $\overline{R}$ minus the arcs of $\alpha$.* (See Fig. 6.)

Let $\epsilon_1$ be a positive number such that if $a_1$ and $b_1$ denote any arcs of $\alpha$ which have no point in common then the distance from any point of $a_1$ to any point of $b_1$ is greater than $\epsilon_1$. Let $\epsilon_2$ be a positive number such that a circle with $P$ as center and $\epsilon_2$ as radius neither contains nor encloses any point of any arc of $\alpha$ or of $ABCD$. Let $k$ be any integer greater than both $1/\epsilon_1$ and $1/\epsilon_2$. Suppose $\beta$ is a double ruling of $R$ with properties (1), (2), (3), and (4) as given above. Let $E$ denote the component, containing $P$, of $\overline{R}$ minus the arcs of $\beta$ which are parallel to $AB$ (for example). In view of property (3) and the additional fact that $1/n<\epsilon_2$ it is obvious that $a_1$ and $a_2$, the arcs of $\beta$ on the boundary of $E$, do not belong to $\alpha$. From (4) it follows that both $a_1$ and $a_2$ have property $c_{1/n}$. Since $1/n<\epsilon_1$ it follows that between each two distinct arcs of $\beta$ parallel to $BC$ the arc $a_i(i=1, 2)$ contains a point not belonging to $M$. Hence if $m+1$ denotes the number of arcs of $\beta$ which are parallel to $BC$ it is easily seen that there exist $2m$ circles lying in $R$ with interiors $C_{11}, C_{12}, \cdots, C_{1m}, C_{21}, C_{22}, \cdots, C_{2m}$ such that (1) $C_{ik}(i=1, 2; k \leqq m)$

contains no point of $M$ and no point of any arc of $\alpha$, and (2) for each two adjacent arcs $b_1$ and $b_2$ of $\alpha$ which are parallel to $BC$ there exist integers $i$ and $j$ such that $C_{1i}$ and $C_{2j}$ contain points of $a_1$ and $a_2$, respectively, which lie between $b_1$ and $b_2$. With the help of Lemmas IV and V it is seen that there exists in $E$ an arc $a_E$ with property $c_{1/(n+1)}$ which together with the arcs of $\beta$ forms a double ruling of $R$, and in addition contains a point in $C_{ik}(i=1, 2; k=1, 2, \cdots, m)$. Obviously no component of $M\cdot(E-a_E)$ contains points in more than two components of $\overline{R}$ minus the arcs of $\alpha$.

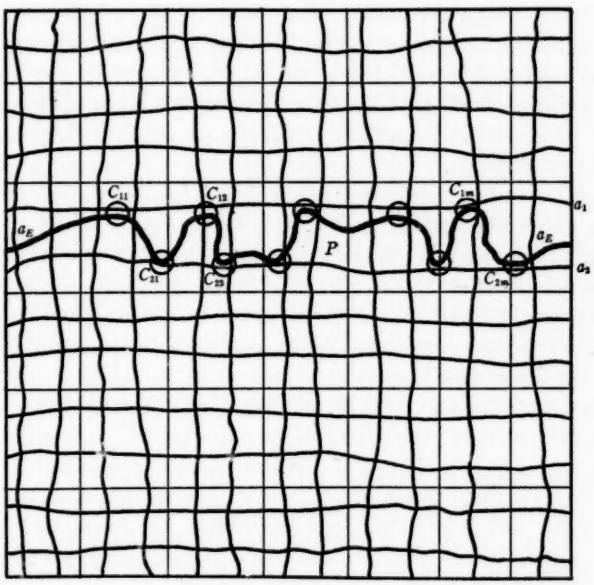


FIG. 6

**Proof of Theorem X.** Suppose $M$ lies in the interior $R$ of a square $ABCD$. With the help of Lemmas IV, V, and VI one can readily see that there exists a double ruling $\beta_1$ of $ABCD$ such that (1) every component of $\overline{R}$ minus the arcs of $\beta_1$ is of diameter less than 1, (2) every arc of $\beta_1$ has property $c_1$, and (3) between each two adjacent arcs of $\beta_1$ parallel to $BC$ there exists on each arc of $\beta_1$ parallel to $AB$ a point not belonging to $M$. Let $a_1$ and $a_2$ denote any two adjacent arcs of $\beta_1$ parallel to $AB$ (or $a_1$ or $a_2$ may be $AB$ or $CD$) and let $E$ denote the set of all points of $\overline{R}$ which lie between $a_1$ and $a_2$. If $m+1$ denotes the number of arcs of $\beta_1$ which are parallel to $BC$ it is clear that

there exist $2m$ circles lying in $R$ with interiors $C_{11}, C_{12}, \cdots, C_{1m}, C_{21}, C_{22}, \cdots, C_{2m}$ such that (1) $C_{ik}(i=1, 2; k \leqq m)$ contains no point of $M$ and no point of any arc of $\beta_1$ which is parallel to $BC$, and (2) for each two adjacent arcs $b_1$ and $b_2$ of $\beta_1$ parallel to $BC$ there exist integers $j$ and $k$ such that $C_{1j}$ and $C_{2k}$ contain points of $a_1$ and $a_2$, respectively, which lie between $b_1$ and $b_2$. With the help of Lemmas IV and V it is seen that there exists in $E$ an arc $a_E$ which together with the arcs of $\beta_1$ forms a double ruling of $R$ and such that (1) $a_E$ contains points in the set $C_{ik}(i=1, 2; k=1, 2, \cdots, m)$, and (2) $a_E$ has property $c_{1/2}$. Obviously no component of $M \cdot (E-a_E)$ contains points in more than two components of $\overline{R}$ minus the arcs of the ruling $\beta_1$. Let $\alpha_1$ be the double ruling obtained by adding to $\beta_1$ the arc $a_E$ for every component $E$ of $\overline{R}$ minus the arcs of $\beta_1$ which are parallel to $AB$. Let $\beta_2$ be a double ruling which contains $\alpha_1$ and is such that (1) every component of $\overline{R}$ minus the arcs of $\beta_2$ is of diameter less than $1/2$ and (2) the arcs of $\beta_2$ which do not belong to $\beta_1$ have property $c_{1/2}$.

Now let $E$ denote the set of all points of $\overline{R}$ which lie between two adjacent arcs of $\beta_2$ which are parallel to $BC$. If there exists in $E$ an arc $a^*$ which together with the arcs of $\beta_2$ forms a double ruling of $R$ and such that (1) no component of $M \cdot (E-a^*)$ has points in more than two components of $\overline{R}$ minus the arcs of $\beta_2$, and (2) the arc $a^*$ has property $c_{1/3}$, then let $a_E$ be such an arc $a^*$. If no such arc exists, but there does exist an arc $a^{**}$ having the above properties except that in (1) the symbol $\beta_1$ replaces the symbol $\beta_2$, then let $a_E$ denote such an arc $a^{**}$. If neither $a^*$ nor $a^{**}$ exists let $a_E$ be any arc in $E$ which together with the arcs of $\beta_2$ forms a double ruling of $R$ and which has property $c_{1/3}$. Let $\alpha_2$ be the double ruling obtained by adding to $\beta_2$ the arc $a_E$ for every component $E$ of $\overline{R}$ minus the arcs of $\beta_2$ which are parallel to $BC$. Let $\beta_3$ be a double ruling which contains $\alpha_2$ and is such that (1) every component of $\overline{R}$ minus the arcs of $\beta_3$ is of diameter less than $1/3$, and (2) the arcs of $\beta_3$ which do not belong to $\beta_2$ have property $c_{1/3}$.

Proceeding in this way one can show that there exists an infinite sequence of double rulings $\beta_1, \beta_2, \beta_3, \cdots$, of $R$ such that for every $n$ the following properties obtain: (1) $\beta_{n+1}$ contains $\beta_n$, (2) every component of $\overline{R}$ minus the arcs of $\beta_n$ is of diameter less than $1/n$, (3) every arc of $\beta_{n+1}$ which does not belong to $\beta_n$ has property $c_{1/(n+1)}$, and (4) if $a_{1n}$ and $a_{2n}$ are adjacent arcs of $\beta_n$ which are parallel to $AB$ for $n$ odd and parallel to $BC$ for $n$ even, and $E$ is the set of all points of $\overline{R}$ which lie between $a_{1n}$ and $a_{2n}$, then if there exists a positive integer $j(j \leqq n-1)$ and an arc $a_{3E}$ which lies in $E$ such that (a) the arc $a_{3E}$ together with the arcs of $\beta_n$ forms a double ruling of $R$, and (b) no component of $M \cdot (E-a_{3E})$ contains points in more than two components of $\overline{R}$ minus the arcs of $\beta_j$, then, $k_E$ denoting the largest such integer $j$, $\beta_{n+1}$

contains an arc having the properties stated above for the arc $a_{jE}$ with $j$ replaced by $k_E$.

Let $P$ denote a point of $R$ not belonging to any arc of any of the double rulings $\beta_1, \beta_2, \beta_3, \cdots$. For each positive integer $n$ let $E_{n,P}$ denote the component, containing $P$, or $\overline{R}$ minus the arcs of $\beta_n$ which are parallel to $AB$ (for example). In view of properties (1), (2), and (3), and Lemma VII, it can be seen that the integer $n$ can be taken large enough so that the integer $j$ as qualified in property (4) does exist for $E_{n,P}$, and furthermore $k_{E_{n,P}}$ increases indefinitely as $n$ increases indefinitely. Hence it follows that for every connected subset $L$ of $M$ there exists an integer $n_L$ such that some arc of $\beta_{n_L}$ parallel to $AB[BC]$ has a point in common with $L$. In view of properties (1) and (2) it follows by methods employed in proving previous theorems that there exists a continuous transformation $T_1$ of $\overline{R}$ into itself such that (1) for every $n$, $T_1$ throws the arcs of $\beta_n$ into intervals of rational lines, and (2) if $L$ is any rational line then there exists an integer $n$ and an arc $g$ of $\beta_n$ such that $T_1(g)$ is a subset of $L$. Obviously there exists a continuous transformation $T$ of the plane $S$ into itself which reduces to $T_1$ for points of $\overline{R}$. Such a transformation satisfies the conclusion of the theorem.

Now as shown in the proof of Theorem IX a continuous curve $M$ which contains a domain contains a continuous curve $M_1$ such that $M_1$ contains no domain but does contain every boundary point of $M$, and such that if $D$ is a complementary domain of $M_1$ which is not a complementary domain of $M$ then $\overline{D}$ lies in a domain of $M$. In view of this fact and the previous theorem the following corollaries may be easily established.

COROLLARY 1. *If $M$ is a bounded continuous curve then there exists a continuous transformation $T$ of the plane into itself such that if $AB$ is an arc and $T(AB)$ is an interval of some rational line then* (1) *$AB \cdot M$ is the sum of a finite number of connected sets, and* (2) *every arc which is a subset of $AB \cdot M$ contains a subarc which either lies on the boundary of a complementary domain of $M$ or lies in a domain which belongs to $M$.*

COROLLARY 2. *If $M$ is a bounded continuous curve and $P$ is a point of $M$ which is not in a domain belonging to $M$ then there exists a continuous transformation $T$ of $S$ into itself such that if $APB$ is any arc such that $T(APB)$ is a subset of a horizontal line, then the component of $M \cdot APB$ which contains $P$ is $P$, and if $AB$ is an arc such that $T(AB)$ is a subset of a rational horizontal line then the number of components of $AB \cdot M$ is finite.*

Let $A$ be any point and for each $n$ let $C_n$ be a circle of radius $1/n$ and center $A$. Let $AB$ be a unit interval and let $M$ be the continuous curve

$AB+C_1+C_2+\cdots$. This example shows that it is not true that if $M$ is any continuous curve then there exists a continuous transformation $T$ of the plane $S$ into itself such that if $AB$ is an arc and $T(AB)$ is a horizontal interval then $AB \cdot M$ is the sum of a finite number of connected sets.

THEOREM XI. *If $P$ is a point of a bounded continuous curve $M$ then there exists an upper semi-continuous collection† $G$ of subcontinua of $M$ which fills up $M$ such that $P$ is an element of $G$ and $G$ is a regular curve with respect to its elements.*

Suppose first that $P$ is a point which does not belong to a domain which belongs to $M$ and let $T$ denote a transformation satisfying the conclusion of Corollary 2. For each point $x$ of $M$ let $g_x$ be the greatest continuum containing $x$ such that $T(g_x)$ is a subset of some horizontal line, and let $G$ denote the collection of continua $g_x$ for all points $x$ of $M$. Clearly $g_P=P$, and the collection $G$ is upper semi-continuous. Now if $M_1$ is any continuum such that the common part of any rational horizontal line and $M_1$ is a finite point set, and the common part of any horizontal line and $M_1$ is totally disconnected, then $M_1$ is a regular curve. Hence $G$ is a regular curve with respect to its elements.

Suppose $P$ is a point lying in a domain $D$ of $M$. There exists a set $K$ of mutually exclusive simple closed curves lying in $D$, all enclosing $P$, no two having a point in common, and such that every point of $D-P$ belongs to some curve of the set $K$. Let $G$ be the upper semi-continuous collection of continua consisting of the curves of the collection $K$ and the continua $M-D$ and $P$. The collection $G$ is an arc with respect to its elements and one of its elements is $P$.

† See R. L. Moore, *Concerning upper semi-continuous collections of continua*, these Transactions, vol. 27 (1925), pp. 416–428. A collection $G$ of continua is said to be an *upper semi-continuous* collection if for each element $g$ of the collection $G$ and each positive number $e$ there exists a positive number $d$ such that if $x$ is any element of $G$ at a lower distance from $g$ less than $d$ then the upper distance of $x$ from $g$ is less than $e$. If $M$ is a point set and $P$ is a point, then by $l(PM)$ is meant the lower bound of the distances from $P$ to all the different points of $M$. If $M$ and $N$ are two point sets, then by $l(MN)$ is meant the lower bound of the values $[l(PN)]$ for all points $P$ of $M$, while by $u(MN)$ is meant the upper bound of these values for all points $P$ of $M$. The point set $M$ is said to be at the upper distance $u(MN)$ from the point set $N$ and is said to be at the lower distance $l(MN)$ from $N$.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

# PAIRS OF RECTILINEAR CONGRUENCES WITH GENERATORS IN ONE-TO-ONE CORRESPONDENCE*

BY

ALEXANDER J. COOK

## 1. Introduction

The place which the configuration composed of a pair of rectilinear congruences with generators in one-to-one correspondence occupies in differential geometry is of some prominence. When corresponding generators do not intersect, a general theory for the projective study of the configuration may be constructed from a system of linear partial homogeneous differential equations of the first order in four dependent and two independent variables.

The theory seems to be well adapted for the study of certain correspondences of Fubini, and a complete basis is furnished for their further investigation. The relation of Fubini's work to the earlier work of Darboux is also fully established. In connection with these correspondences a generalization to surfaces is secured for the concept of intersector curve which Lane introduced in his study of pairs of ruled surfaces.

The consideration of corresponding nets of curves on the intersector surfaces suggests itself, and when the nets of curves are conjugate nets on the surfaces, two distinct possibilities occur. The first of these has been studied by Fubini, and the properties which he discovered are readily obtained by the methods of the present theory. In addition, one or two new theorems are discovered. The second case arises when a family of intersector surfaces exists in association with each of the two congruences in correspondence.

## 2. The differential equations

Let

$$y_k^i = y_k^i(u,v) \qquad (k = 1,2,3,4)$$

be the equations of four† surfaces $S_i (i=1, \cdots, 4)$. When $u=u_0$, $v=v_0$, four points $P_i$ are obtained, one for each surface. Let $l_1$ be the line joining $P_1$, $P_2$, and $l_2$ the join of $P_3$, $P_4$. As $u$ and $v$ take on all admissible values,

* Presented to the Society, June 20, 1929; received by the editors in April, 1929.

† The subscript or superscript letters $h, i, j, k$ will be taken to mean (1, 2, 3, 4) throughout the discussion.

$l_1$ generates a rectilinear congruence $\Gamma_1$ as long as $S_1$ and $S_2$ do not degenerate into fixed points, into curves whose parameters are dependent, or finally into a fixed point and a curve respectively. Likewise the lines $l_2$ ordinarily determine a congruence $\Gamma_2$, whose generators are in one-to-one correspondence with the generators of $\Gamma_1$.

It will be assumed in what follows that $l_1$ and $l_2$ are skew lines, so that

$$\Delta = | y^1 y^2 y^3 y^4 | \neq 0.$$

A system of linear differential equations of the first order may therefore be constructed of the form

$$y_u^i = a^{ij} y^j, \quad y_v^i = b^{ij} y^j, \tag{2.1}$$

where each $y^i$ stands for $y_k^i$, and where the summation convention used in tensor analysis has been introduced.

Suppose now that we are presented with a system of equations in four dependent variables $y^i$ and two independent variables $u$ and $v$ of the form (2.1). The system will be completely integrable when and only when the conditions

$$a_v^{ij} + a^{ik} b^{kj} = b_u^{ij} + b^{ik} a^{kj} \tag{2.2}$$

are satisfied. In particular we find

$$\sum_{i=1}^{4} a_v^{ii} = \sum_{i=1}^{4} b_u^{ii}. \tag{2.3}$$

Hence $f(u, v)$ exists such that

$$f_u = \sum_{i=1}^{4} a^{ii}, \quad f_v = \sum_{i=1}^{4} b^{ii},$$

and further

$$\Delta = ce^f, \quad c = \text{const.}$$

The most general system of solutions of (2.1) is a linear transform, with constant coefficients, of any particular fundamental system of solutions $y_k^i$. The equations (2.1) and (2.2) thus define four surfaces $S_i$, which taken in pairs determine six congruences; these in turn can be grouped in twos, and in any pair corresponding generators do not intersect. The congruences $\Gamma_1$ and $\Gamma_2$ form such a pair and the study of the correspondence existing between them by means of (2.1) and (2.2) will necessarily be a projective study.

The geometry of the correspondence between $\Gamma_1$ and $\Gamma_2$ is equivalent to the theory of the invariants and covariants of (2.1) under the most general transformation

$$(2.4)\quad \begin{aligned} &y^m = \alpha_m \bar{y}^1 + \beta_m \bar{y}^2, \qquad y^n = \alpha_n \bar{y}^3 + \beta_n \bar{y}^4 \qquad (m = 1,2\,;\ n = 3,4),\\ &\Delta_{12} = \alpha_1\beta_2 - \alpha_2\beta_1 \neq 0, \quad \Delta_{34} = \alpha_3\beta_4 - \alpha_4\beta_3 \neq 0, \end{aligned}$$

and

$$(2.5)\qquad \bar{u} = \phi(u,v), \quad \bar{v} = \psi(u,v), \quad J(\phi,\psi) \neq 0,$$

which leaves the two congruences unaltered.* The system (2.1) under (2.4) goes over into a system of the same form whose coefficients are readily computed. It is only necessary to display eight of these, namely

$$(2.6)\quad \begin{aligned} \Delta_{12}\bar{a}^{11} &= \beta_2(-\alpha_{1u} + a^{1m}\alpha_m) - \beta_1(-\alpha_{2u} + a^{2m}\alpha_m),\\ \Delta_{12}\bar{a}^{12} &= \beta_2(-\beta_{1u} + a^{1m}\beta_m) - \beta_1(-\beta_{2u} + a^{2m}\beta_m),\\ \Delta_{12}\bar{a}^{21} &= -\alpha_2(-\alpha_{1u} + a^{1m}\alpha_m) + \alpha_1(-\alpha_{2u} + a^{2m}\alpha_m),\\ \Delta_{12}\bar{a}^{22} &= -\alpha_2(-\beta_{1u} + a^{2m}\beta_m) + \alpha_1(-\beta_{2u} + a^{2m}\beta_m),\\ \Delta_{12}\bar{a}^{13} &= \beta_2 a^{1n}\alpha_n - \beta_1 a^{2n}\alpha_n, \quad \Delta_{12}\bar{a}^{23} = -\alpha_2 a^{1n}\alpha_n + \alpha_1 a^{2n}\alpha_n,\\ \Delta_{12}\bar{a}^{14} &= \beta_2 a^{1n}\beta_n - \beta_1 a^{2n}\beta_n, \quad \Delta_{12}\bar{a}^{24} = -\alpha_2 a^{1n}\beta_n + \alpha_1 a^{2n}\beta_n. \end{aligned}$$

Another eight coefficients are secured from the substitution $(ab)$ $(uv)$ and the remaining sixteen are obtained by applying the substitution

$$(2.7)\qquad (13)(24)(mn)$$

to those already determined.

### 3. A CANONICAL FORM FOR THE DIFFERENTIAL EQUATIONS

New reference surfaces for the congruences may be introduced by applying (2.4) to (2.1). Let $\alpha_l$, $\beta_l$ be arbitrary pairs of solutions of the equations

$$-\alpha_{lu} + a^{lm}\alpha_m = 0, \qquad -\beta_{lv} + b^{lm}\beta_m = 0,$$

so chosen that $\Delta_{12} \neq 0$. A glance at (2.6) gives the relations

$$(3.1_1)\qquad \bar{a}^{11} = \bar{a}^{21} = \bar{b}^{12} = \bar{b}^{22} = 0,$$

and the reference surfaces of $\Gamma_2$ may be chosen similarly to give the relations

$$(3.1_2)\qquad \bar{a}^{33} = \bar{a}^{43} = \bar{b}^{34} = \bar{b}^{44} = 0.$$

The equations of the correspondence now take the form (2.1) where the conditions (3.1) are satisfied. The most general transformation (2.4) and (2.5) which preserves (3.1) and leaves the reference surfaces unaltered is of the form

* The meaning here given to the letters $m$ and $n$ will be preserved throughout the discussion. The letters $l$ and $p$ will be used in the sense of $m$ and $n$ respectively.

$$y^1 = \alpha_1(v)\bar{y}^1, \quad y^2 = \beta_2(u)\bar{y}^2, \quad y^3 = \alpha_3(v)\bar{y}^3, \quad y^4 = \beta_4(u)\bar{y}^4,$$
$$\bar{u} = \phi(u), \quad \bar{v} = \psi(v).$$

## 4. Pairs of congruences with an intersector property

In his study of pairs of ruled surfaces, Lane* has shown that a family of curves, which he has called a family of intersector curves, exists on each ruled surface, the curves being defined as follows. A curve on a ruled surface $R_1$ is an intersector curve with respect to a second ruled surface $R_2$ in case the tangent at each point of the curve intersects the line $l_2$ which corresponds to the generator $l_1$ that passes through the point. We now proceed to describe a somewhat analogous situation for a pair of congruences. Let $\Gamma_1$ be given together with a one-parameter family of surfaces so that $l_1$ cuts each of the surfaces in just one point. Further let us suppose that the tangent planes to the surfaces at the points of $l_1$ form a pencil; then the axes of the pencils so obtained determine $\Gamma_2$, which, by construction, has its generators in one-to-one correspondence with the generators† of $\Gamma_1$. The surfaces of the family will be called *intersector surfaces* of $\Gamma_1$ with respect to $\Gamma_2$, and we shall say that $\Gamma_1$ possesses *property I* with respect to $\Gamma_2$.

The conditions under which the reference surfaces of $\Gamma_1$ can be chosen to have the intersector property with respect to $\Gamma_2$ are readily determined. For convenience, let $\alpha_1 = -\lambda_1\alpha_2$, $\beta_1 = -\lambda_2\beta_2$, where $\lambda_1 \neq \lambda_2$, since $\Delta_{12} \neq 0$. Then $\lambda_1$ and $\lambda_2$ appear as distinct solutions of the equations

$$\begin{aligned} \lambda_u &= a^{21}\lambda^2 + (a^{11} - a^{22})\lambda - a^{12}, \\ \lambda_v &= b^{21}\lambda^2 + (b^{11} - b^{22})\lambda - b^{12}, \end{aligned} \tag{4.1}$$

which can have analytic solutions only when the equation

$$p\lambda^2 + q\lambda + r = 0, \quad D^2 = q^2 - 4pr, \tag{4.2}$$

is satisfied, where

$$\begin{aligned} p &= a_v^{21} - b_u^{21} + a^{2m}b^{m1} - b^{2m}a^{m1}, \\ q &= a_v^{11} - b_u^{11} - (a_v^{22} - b_u^{22}) + a^{1m}b^{m1} - b^{1m}a^{m1} - (a^{2m}b^{m2} - b^{2m}a^{m2}), \\ r &= a_v^{12} - b_u^{12} + a^{1m}b^{m2} - b^{1m}a^{m2}. \end{aligned} \tag{4.3}$$

---

* E. P. Lane, *Ruled surfaces with generators in one-to-one correspondence*, these Transactions, vol. 25 (1923), pp. 287–290.

† Correspondences of this type have been discussed by G. Fubini, *Su alcune classi di congruenze di rette e sulle transformazione delle superficie R*, Annali di Matematica, (4), vol. 1 (1923–24), pp. 241–257.

The expressions for $p, q, r$ may be written as

$$
\begin{aligned}
p &= -a^{2n}b^{n1} + b^{2n}a^{n1}, \\
q &= -(a^{1n}b^{n1} - b^{1n}a^{n1}) + (a^{2n}b^{n2} - b^{2n}a^{n2}), \\
r &= -a^{1n}b^{n2} + b^{1n}a^{n2},
\end{aligned} \tag{4.4}
$$

by virtue of the conditions (2.2).

In case $p \equiv q \equiv r \equiv 0$, the pair of equations (4.1) will be spoken of as a *Riccati system* of differential equations, with the integrability conditions

$$
\begin{aligned}
a^{1n}b^{n2} &= b^{1n}a^{n2}, \\
a^{1n}b^{n1} - a^{2n}b^{n2} &= b^{1n}a^{n1} - b^{2n}a^{n2}, \\
a^{2n}b^{n1} &= b^{2n}a^{n1}.
\end{aligned} \tag{4.5}
$$

The conditions (4.5) are necessary and sufficient for the existence of a one-parameter set of solutions $\lambda$ of the Riccati system (4.1), hence the theorem. *The conditions* (4.5) *on the coefficients of* (2.1) *provide a necessary and sufficient condition that* $\Gamma_1$ *possess property I with respect to* $\Gamma_2$. *Furthermore, the intersector surfaces of* $\Gamma_1$ *with respect to* $\Gamma_2$ *are determined by solving a Riccati system.*

The intersector surfaces are in fact defined by

$$x = y^1 + \lambda y^2, \tag{4.6}$$

where $\lambda = \lambda(u, v)$ is an arbitrary solution of (4.1).

Necessary and sufficient conditions that $\Gamma_2$ possess property $I$ with respect to $\Gamma_1$ are obtained by applying (2.7) to (4.5). These conditions are not a consequence of (4.5), so that the two sets of conditions taken together provide a necessary and sufficient condition that $\Gamma_1$ and $\Gamma_2$ possess property $I$ with respect to each other.

There remain for discussion the solutions of (4.1) when not all of $p, q, r$ vanish identically. In this case the analytic solutions of (4.1) are not more than two in number; for, if $\lambda = \lambda(u, v)$ is an analytic solution of (4.1) it is thereby a solution of (4.2). Suppose conversely that $\lambda$ is a solution of (4.2); the conditions under which $\lambda$ is a solution of (4.1) are the following:

$$
\begin{aligned}
pq_u - p_u q &= 2p(a^{12}p + a^{21}r) + q(-a^{21}q + (a^{11} - a^{22})p), \\
pr_u - p_u r &= q(a^{12}p + a^{21}r) + 2r(-a^{21}q + (a^{11} - a^{22})p), \\
pq_v - p_v q &= 2p(b^{12}p + b^{21}r) + q(-b^{21}q + (b^{11} - b^{22})p), \\
pr_v - p_v r &= q(b^{12}p + b^{21}r) + 2r(-b^{21}q + (b^{11} - b^{22})p).
\end{aligned} \tag{4.7}
$$

*In case* $D \neq 0$ *the conditions* (4.7) *are necessary and sufficient for the existence*

*of two and only two surfaces associated with* $\Gamma_1$ *having the intersector property with respect to* $\Gamma_2$.

The two surfaces are given by (4.6) where $\lambda$ satisfies (4.2). They may be chosen as reference surfaces for $\Gamma_1$ in which case we have the relations

$$a^{12} = a^{21} = b^{12} = b^{21} = 0. \tag{4.8}$$

When $D=0$ the conditions (4.7) reduce to the form

$$\begin{vmatrix} pq_u - p_u q, & pr_u - p_u r \\ pq_v - p_v q, & pr_v - p_v r \end{vmatrix} = 0. \tag{4.9}$$

*In case* $D=0$, *and* $p, q, r$ *are not all identically zero,* (4.9) *provides a necessary and sufficient condition for the existence of one and only one surface associated with* $\Gamma_1$ *having the intersector property with respect to* $\Gamma_2$. As before, the surface in question is given by (4.6) and (4.2). The surface may be chosen as the reference surface $S_1$ for $\Gamma_1$ in which case $a^{12}=b^{12}=0$.

## 5. The theory of a Riccati system of differential equations

Let there be given a Riccati system

$$\begin{aligned} \lambda_u &= a_1\lambda^2 + a_2\lambda + a_3, \\ \lambda_v &= b_1\lambda^2 + b_2\lambda + b_3, \end{aligned} \tag{5.1}$$

where $a_r, b_r$ $(r=1, 2, 3)$ are analytic functions of $u$ and $v$, with the integrability conditions

$$\begin{aligned} a_{1v} - b_{1u} + (a_1b_2 - a_2b_1) &= 0, \\ a_{2v} - b_{2u} + 2(a_1b_3 - a_3b_1) &= 0, \\ a_{3v} - b_{3u} + (a_2b_3 - a_3b_2) &= 0. \end{aligned} \tag{5.2}$$

Theorems analogous to those for an ordinary Riccati equation* may be obtained for a Riccati system (5.1). Thus, for instance, *the cross ratio of any four analytic solutions of a Riccati system is constant.*

Let us now return to the geometrical considerations from which arose the discussion of the Riccati system of equations. In the first place the meaning of the term intersector surface becomes apparent from the following theorem. *When* $\Gamma_1$ *possesses property I with respect to* $\Gamma_2$, *the intersector curves on all the ruled surfaces* $R_1$ *of* $\Gamma_1$ *with respect to the corresponding ruled surfaces* $R_2$ *of* $\Gamma_2$ *form themselves into the intersector surfaces of* $\Gamma_1$ *with respect to* $\Gamma_2$. *This theorem can be phrased more directly by saying that any ruled surface*

* For a discussion of this equation see L. P. Eisenhart, *Differential Geometry*, Boston, Ginn, 1909, pp. 25–26.

*$R_1$ of $\Gamma_1$ meets the one-parameter family of intersector surfaces along its intersector curves with respect to the corresponding ruled surface $R_2$ of $\Gamma_2$.* In this form the theorem is geometrically evident and we can in fact exhibit the Riccati equation defining the intersector curves of $R_1$. The parameters $u$ and $v$ can always be chosen so that any $R_1$ of $\Gamma_1$ is given by $v=v(u)$ which also defines $R_2$. We have from (2.1) the relations

$$\frac{dy^i}{du} = (a^{ij} + v'b^{ij})y^j, \tag{5.3}$$

which, on comparison with Lane, loc. cit., p. 282, (1), are the equations for the correspondence between $R_1$ and $R_2$, and on the other hand,

$$\frac{d\lambda}{du} = (a^{21} + v'b^{21})\lambda^2 + \{(a^{11} - a^{22}) + v'(b^{11} - b^{22})\}\lambda - (a^{12} + v'b^{12}),$$

which, from (5.3) and Lane, loc. cit., p. 288, (13), is the required Riccati equation.

The cross ratio of any four solutions of a Riccati system having been shown to be constant, we have proved that *when $\Gamma_1$ possesses property $I$ with respect to $\Gamma_2$, any four intersector surfaces cut each generator $l_1$ of $\Gamma_1$ in a set of four points which have the same cross ratio on all generators of $\Gamma_1$.* This theorem is a generalization of the theorem proved by Lane, loc. cit., p. 289, for intersector curves.

Since the theory of a Riccati equation is equivalent to the theory of the intersector curves on a given ruled surface, with respect to some associated ruled surface,* it is natural to inquire if there is any sense in which it can be said that the theory of a Riccati system is equivalent to the theory of the intersector surfaces of a given congruence which possesses property $I$ with respect to some associated congruence. The answer is in the negative, for let us assume that we are given a Riccati system (5.1) and a congruence $\Gamma_1$; then it is not possible to determine a second congruence $\Gamma_2$ with respect to which $\Gamma_1$ possesses property $I$, and at the same time have the intersector surfaces defined by (5.1). We omit the proof of this statement.

If the parametric ruled surfaces of the congruence $\Gamma_1$ for the correspondence (2.1) are chosen as the developables of $\Gamma_1$ the relations

$$a^{13}a^{24} - a^{14}a^{23} = 0, \quad b^{13}b^{24} - b^{14}b^{23} = 0 \tag{5.4}$$

are obtained, evidently necessary and sufficient conditions that the developables of $\Gamma_1$ be the parametric surfaces for that congruence in the corre-

* Lane, loc. cit., p. 289.

pondence defined by (2.1). The application of the substitution (13) (24) to (5.4) yields the relations

$$a^{31}a^{42} - a^{32}a^{41} = 0, \quad b^{31}b^{42} - b^{32}b^{41} = 0 \tag{5.5}$$

which are necessary and sufficient that the developables of $\Gamma_2$ be the parametric surfaces for $\Gamma_2$.

## 6. The intersector surfaces

When $\Gamma_1$ possesses property $I$ with respect to $\Gamma_2$ the equations of the correspondence and of the intersector surfaces reduce to simpler forms. The intersector surfaces form a family, and the points of any pair of the surfaces are in correspondence, where corresponding points lie on a generator $l_1$ of $\Gamma_1$. The consideration of corresponding nets of curves on these surfaces suggests itself. When the nets of curves are conjugate nets two distinct possibilities occur. The first of these, in which case the developables of $\Gamma_1$ trace out conjugate nets on all the intersector surfaces, has been studied by Fubini* who at the same time recognized† its relation to certain theorems of Darboux.‡ This relation may be fully established by means of the present theory. The second case occurs when each of the congruences $\Gamma_1$ and $\Gamma_2$ possesses property $I$ with respect to the other.

When $\Gamma_1$ possesses property $I$ with respect to $\Gamma_2$ the conditions (4.5) are satisfied, and by choosing an arbitrary pair of intersector surfaces as reference surfaces for $\Gamma_1$, the coefficients of (2.1) can be made to satisfy the additional relations

$$a^{11} = a^{12} = a^{21} = a^{22} = b^{12} = b^{21} = 0, \quad b^{11} = b^{22}. \tag{6.1}$$

The intersector surfaces are then given by (4.6), where $\lambda = \text{const.}$ By means of (4.6) the coefficients $D$, $D'$, $D''$ may be computed for an arbitrary intersector surface $S_x$; they are

$$\begin{aligned} D &= (a^{1n}a^{n2} + \lambda(a^{2n}a^{n2} - a^{1n}a^{n1}) - \lambda^2 a^{2n}a^{n1})F\Delta, \\ D' &= (a^{1n}b^{n2} + \lambda(a^{2n}b^{n2} - a^{1n}b^{n1}) - \lambda^2 a^{2n}b^{n1})F\Delta, \\ D'' &= (b^{1n}b^{n2} + \lambda(b^{2n}b^{n2} - b^{1n}b^{n1}) - \lambda^2 b^{2n}b^{n1})F\Delta, \end{aligned} \tag{6.2}$$

where

$$F = \begin{vmatrix} a^{13} + \lambda a^{23}, & a^{14} + \lambda a^{24} \\ b^{13} + \lambda b^{23}, & b^{14} + \lambda b^{24} \end{vmatrix}. \tag{6.3}$$

---

* Fubini, loc. cit., pp. 246–250.

† Ibid., p. 242.

‡ G. Darboux, *Leçons sur la Théorie Générale des Surfaces*, 1915, 2e édition, livre IV, chap. X, pp. 241–252.

When $F=0$, (4.6) defines the focal surfaces of $\Gamma_1$. Evidently a focal surface cannot have the intersector property; the focal surfaces must also be distinct, otherwise the intersector surfaces degenerate into intersector curves, a situation which will be excluded.

The equations (6.2) show that the parametric surfaces of $\Gamma_1$ will trace out a conjugate net of curves on every intersector surface when the following conditions are satisfied:

$$a^{1n}b^{n2}=0,\quad a^{2n}b^{n2}=a^{1n}b^{n1},\quad a^{2n}b^{n1}=0\ ; \tag{6.4}$$

from (4.5) the further relations

$$b^{1n}a^{n2}=0,\quad b^{2n}a^{n2}=b^{1n}a^{n1},\quad b^{2n}a^{n1}=0 \tag{6.5}$$

are obtained. The conditions $(6.4_1)$, $(6.4_3)$ and $(6.5_1)$, $(6.5_3)$ result in the equations

$$a^{2n}b^{n2}=\frac{a^{24}b^{42}}{a^{13}b^{31}}a^{1n}b^{n1},\quad b^{2n}a^{n2}=\frac{b^{24}a^{42}}{b^{13}a^{31}}b^{1n}a^{n1}.$$

A comparison of these two relations with $(6.4_2)$ and $(6.5_2)$ leads to three possible consequences of (6.4) and (6.5), namely:

$$a^{1n}b^{n1}=0,\qquad b^{1n}a^{n1}=0, \tag{6.6}$$

$$a^{24}b^{42}-a^{13}b^{31}=0,\quad b^{24}a^{42}-b^{13}a^{31}=0, \tag{6.7}$$

and

$$a^{1n}b^{n1}=0,\quad b^{24}a^{42}-b^{13}a^{31}=0, \tag{6.8}$$

or the alternatives

$$b^{1n}a^{n1}=0,\quad a^{24}b^{42}-a^{13}b^{31}=0, \tag{6.9}$$

which are not essentially different from (6.8).

We turn attention to the situation corresponding to (6.4), (6.5), (6.6). These conditions taken together may be written in the form

$$a^{1n}b^{nm}=0,\quad b^{1n}a^{nm}=0. \tag{6.10}$$

As a consequence we note first of all that (5.4) and (5.5) are evidently satisfied. Geometrically, (6.10) then implies that the developables of $\Gamma_1$ are parametric, and the corresponding parametric surfaces for $\Gamma_2$ are the developables of that congruence. The developables of the pair of congruences thus correspond, and those of $\Gamma_1$ trace out a conjugate net of curves on every intersector surface. Secondly, we see on comparing (2.2) and (6.1) with (6.10) that

$$b_u^{11}=0. \tag{6.11}$$

In case $\Gamma_1$ possesses property $I$ with respect to $\Gamma_2$ with the further property that the developables of $\Gamma_1$ trace out a conjugate net of curves on every intersector surface, we shall say for brevity that $\Gamma_1$ possesses *property* $I_1$ with respect to $\Gamma_2$.*

The following two theorems concerning necessary and sufficient conditions for property $I_1$ are stated without proof.

*Let* $\Gamma_1$ *possess property* $I$ *with respect to* $\Gamma_2$ *and write the equations of the correspondence in the form* (2.1), (6.1). *The equation* (6.11) *is a necessary and sufficient condition that* $\Gamma_1$ *possess property* $I_1$ *with respect to* $\Gamma_2$.

*When the equations of the correspondence between* $\Gamma_1$ *and* $\Gamma_2$ *are in general form* (2.1), *the conditions*

$$a^{ln}b^{nm} = b^{ln}a^{nm} \tag{6.12}$$

*are necessary and sufficient that* $\Gamma_1$ *possess property* $I_1$ *with respect to* $\Gamma_2$. *In case the developables of* $\Gamma_1$ *are the parametric surfaces of the congruence these conditions take the form* (6.10).

As a consequence of these results two theorems due to Fubini† may be proved without difficulty:

*When* $\Gamma_1$ *possesses property* $I$ *with respect to* $\Gamma_2$ *the developables of* $\Gamma_1$ *cannot trace out a conjugate net on any of the intersector surfaces unless* $\Gamma_1$ *possess property* $I_1$ *with respect to* $\Gamma_2$.

*When* $\Gamma_1$ *possesses property* $I_1$ *with respect to* $\Gamma_2$, *the developables of the congruences correspond.*

We are now in a position to connect up the fundamental theorem of Darboux‡ with the present theory. Let two surfaces $S_1$ and $S_2$ be given in one-to-one point correspondence. The lines joining corresponding points generate a congruence $\Gamma_1$, while the lines of intersection of the tangent planes to the surfaces at corresponding points generate a second congruence $\Gamma_2$. Corresponding generators do not intersect unless $S_1$ and $S_2$ are the focal surfaces of $\Gamma_1$, in which case $l_1$ and $l_2$ coincide. This case is excluded from the discussion.

When the developables of $\Gamma_1$ trace out conjugate nets on each of the surfaces $S_1$ and $S_2$ the surfaces are said by some writers to be in the relation of a transformation $F$, or briefly in *relation* $F$.§

* This correspondence was first studied by Fubini, loc. cit., pp. 246–250.

† Ibid., pp. 244 and 246.

‡ Darboux, loc. cit., p. 244.

§ L. P. Eisenhart, *Transformations of Surfaces*, Princeton University Press, 1923, p. 34.

We can now prove the following theorem:

*$S_1$ and $S_2$ are in relation $F$ if and only if $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$.*

To show this we note to begin with that (4.7) is necessarily satisfied. As a consequence, the additional relations

$$a^{11} = a^{12} = a^{21} = b^{12} = b^{21} = b^{22} = 0 \tag{6.13}$$

can be made to hold. Further, the developables of $\Gamma_1$ may be chosen as the parametric surfaces of the congruence, so that the conditions (5.4) are satisfied. $S_1$ and $S_2$ are now the reference surfaces for $\Gamma_1$, and the developables of $\Gamma_1$ will trace out a conjugate net of curves on each side of these surfaces only when

$$a^{1n}b^{n2} = 0, \qquad a^{2n}b^{n1} = 0,$$

$S_1$ and $S_2$ being non-degenerate. The conditions (2.2) give the further relations

$$b^{1n}a^{n2} = 0, \qquad b^{2n}a^{n1} = 0.$$

The equations (5.4) may now be brought into play and (6.10) follows thus showing that $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$.

Conversely, if $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$, $S_1$ and $S_2$ are evidently intersector surfaces and the developables of $\Gamma_1$ trace out a conjugate net of curves on each of the surfaces.

Let us turn next to the consideration of (6.4), (6.5) and (6.7). The conditions found from (4.5) by use of (2.7) can be shown to be satisfied so that *$\Gamma_2$ necessarily possesses property $I$ with respect to $\Gamma_1$, and the parametric surfaces of $\Gamma_2$ trace out a conjugate net of curves on every intersector surface of $\Gamma_2$ with respect to $\Gamma_1$.*

Suppose now, conversely, we assume that each of $\Gamma_1$, $\Gamma_2$ possesses property $I$ with respect to the other. The equations of the correspondence can be written in the form (2.1), with (6.1) and with the additional conditions

$$b^{np} = 0, \quad a^{33} = a^{44}, \quad a^{34} = a^{43} = 0. \tag{6.14}$$

We note also from (2.3) that the equation

$$a_v^{33} = b_u^{11} \tag{6.15}$$

is satisfied. The coefficients of this reduced system of equations satisfy (4.5), and the conditions found from (4.5) and (2.7) by the nature of the correspondence which they represent.

Suppose further that the coefficients of (2.1) satisfy (6.7). The con-

ditions (6.7) are equivalent to a particular choice of the parametric ruled surfaces for the correspondence. Then it may be shown that (6.4), (6.5) hold provided the condition

$$a^{24}a^{42} - a^{13}a^{31} \neq 0 \tag{6.16}$$

be maintained.

We have thus proved the following theorem.

*When each of* $\Gamma_1$, $\Gamma_2$ *possesses property I with respect to the other there exists ordinarily one and just one net of ruled surfaces for each congruence which traces out a conjugate net of curves on every intersector surface associated with that congruence.*

The third possibility for consideration is expressed by (6.4), (6.5), (6.8). We note that (6.8) is made up of $(6.6_1)$ and $(6.7_2)$. We might expect interaction between $(6.6_1)$ and $(6.7_2)$, creating conditions not already found in the discussion of the previous cases; the nature of $(6.6_1)$ and $(6.7_2)$, however, shows that this is not possible. Hence we need not expect a third type of relation between $\Gamma_1$ and $\Gamma_2$ for which conjugate nets of curves on the intersector surfaces of $\Gamma_1$ with respect to $\Gamma_2$ are in correspondence.

If the parametric surfaces of $\Gamma_1$ are to trace out an asymptotic net of curves on every intersector surface of $\Gamma_1$ with respect to $\Gamma_2$, three possibilities are secured as in the discussion for conjugate nets of curves.

The conditions secured in the first case cannot hold, for they imply that the developables are the parametric surfaces for $\Gamma_1$, and Fubini* has shown that the developables of $\Gamma_1$ cannot trace out the asymptotic nets on all the intersector surfaces.

In the second case the relations

$$\begin{aligned} &a^{4m}a^{m4} = a^{3m}a^{m3}, \quad a^{nm}a^{mp} = 0, \\ &b^{4m}b^{m4} = b^{3m}b^{m3}, \quad b^{nm}b^{mp} = 0 \qquad (n \neq p) \end{aligned} \tag{6.17}$$

may be obtained in a manner entirely analogous to the discussion of the equations (6.4), (6.5), (6.7). We are unable to say anything further about this case, other than that it does not appear likely to yield conditions sufficient for the desired correspondence of the asymptotic nets on the intersector surfaces.

## 7. The transformation of Laplace

When $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$, the system (2.1) may be simplified so that the relations

* Fubini, loc. cit., p. 246.

$$a^{lm} = b^{lm} = 0 \tag{7.1}$$

are satisfied.

Let the developables of $\Gamma_1$ be chosen as the parametric surfaces for the congruence. The equations (5.4) and (5.5) are satisfied since the developables of the congruences correspond, and (6.12) becomes (6.10). From (5.4) and (7.1) we have the equations

$$y_u^1 = t_1 y_u^2, \quad y_v^1 = t_2 y_v^2, \tag{7.2}$$

where

$$t_1 = a^{13}/a^{23} = a^{14}/a^{24}, \; t_2 = b^{13}/b^{23} = b^{14}/b^{24}, \; t_1 \neq t_2.$$

A generator $l_1$ of $\Gamma_1$ is cut by an arbitrary intersector surface in the point $P_x$ whose coördinates are given by (4.6) with $\lambda = \text{const}$. From (4.6) and (7.2) we find

$$x_u = (t_1 + \lambda) y_u^2, \quad x_v = (t_2 + \lambda) y_v^2,$$

and certain theorems concerning the nature of the correspondence, proved by Fubini* and Darboux,† now follow without difficulty.

Since the developables of $\Gamma_1$ meet any intersector surface in a conjugate net of curves, these curves form the parametric net for the surface, where it is assumed that (5.4) and (5.5) are in force, and the Laplace equation for the surface may be determined as

$$x_{uv} = \frac{t_2 + \lambda}{t_1 + \lambda} b x_u + \frac{t_1 + \lambda}{t_2 + \lambda} c x_v, \tag{7.3}$$

where

$$b = \frac{-t_{1v}}{t_1 - t_2}, \; c = \frac{t_{2u}}{t_1 - t_2}.$$

In case $b = 0$ or $c = 0$, one or the other of the focal surfaces of $\Gamma_2$ degenerates into a curve, so that we may assume $bc \neq 0$.

The following theorem may now be stated without proof.

*The locus of the ray-points $P_\sigma$ of the u-curves of the intersector surfaces at the points of a generator $l_1$ of $\Gamma_1$ is the v-tangent to the focal surface for the u-developables at the focal point of $l_1$.*

*The focal point of $l_1$ is a singular point for the locus.*

A corresponding theorem may be stated by interchanging $u$ and $v$, and replacing $\sigma$ by $\rho$.

* Fubini, loc. cit., pp. 246 and 247.

† Darboux, loc. cit., p. 245. The theorem of Darboux appears as a special case of the theorem of Fubini, loc. cit., p. 247, by virtue of the theorem which has been proved in § 6.

Suppose now we fix upon a particular intersector surface $x=y^1+\lambda_0 y^2$, where $\lambda=\lambda_0$ is some constant. For this surface we find

$$x_u = (t_1 + \lambda_0) y_u^2, \quad t_1 = t_1(u,v).$$

The equation $t_1(u, v)+\lambda_0=0$ defines a certain curve on the surface and $x_u=0$ all along this curve. Every point of this curve is a point $P_{\tau^1}$, its coordinates having the form $\tau^1=y^1-t_1y^2$, hence the curve also lies on the focal surface of the $u$-developable of $\Gamma_1$. At the same time we find

$$x_v = (t_2 + \lambda_0) y_v^2, \quad t_2 = t_2(u,v),$$

so that $x_v=0$ all along the curve defined by $t_2(u, v)+\lambda_0=0$, which is the intersection of the given surface and the focal surface for the $v$-developables of $\Gamma_1$. The following theorem has thus been proved.

*When $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$ every intersector surface of $\Gamma_1$ with respect to $\Gamma_2$ contains two singular curves which are the intersections of the surface with the focal surfaces of $\Gamma_1$.*

The ray-points for the reference surface $S_1$ at $P_{y^1}$ may be written as $P_{\rho^1}, P_{\sigma^1}$, and likewise the ray-points of $S_2$ at $P_{y^2}$ as $P_{\rho^2}, P_{\sigma^2}$ respectively. The line $l_{\rho^1\rho^2}$ is the locus of the ray-point $P_\rho$ and $l_{\sigma^1\sigma^2}$ is the locus of the ray-point $P_\sigma$. Two lines have thus been determined which correspond to $l_1$, thus generating the pair of congruences $\Gamma_\rho, \Gamma_\sigma$.

The Laplace transforms of $\Gamma_1$ are defined by means of the focal surfaces of $\Gamma_1$, whose equations are

$$\tau^1 = y^1 - t_1y^2, \quad \tau^2 = y^1 - t_2y^2,$$

respectively. The congruence $\Gamma_{\tau^2\tau^2_v}$ is the congruence $\Gamma_1$, but $\Gamma_{\tau^2\tau^2_u}$ which, by definition, is called the first Laplace transform of $\Gamma_1$, has been shown to be $\Gamma_\rho$. The congruence $\Gamma_{\tau^1\tau^1_u}$ is again $\Gamma_1$, but $\Gamma_{\tau^1\tau^1_v}$ which, by definition, is the minus first Laplace transform of $\Gamma_1$, is $\Gamma_\sigma$.

The points $P_{y^1_v}, P_{y^1_u}$ are the focal points for $l_2$, and $\Gamma_{y^1_u y^1_{uv}}$ is $\Gamma_2$ while $\Gamma_{y^1_u y^1_{uu}}$ is the first Laplace transform of $\Gamma_2$. $\Gamma_{y^1_v y^1_{uv}}$ is again $\Gamma_2$, while $\Gamma_{y^1_v y^1_{vv}}$ is the minus first Laplace transform of $\Gamma_2$.

The equations (2.1) may be set up for the correspondence between the first Laplace transforms of $\Gamma_1$ and $\Gamma_2$ and the theorem follows:

*When $\Gamma_1$ possesses property $I_1$ with respect to $\Gamma_2$, the first Laplace transform of $\Gamma_1$ possesses property $I_1$ with respect to the first Laplace transform of $\Gamma_2$.* The intersector surfaces for the derived correspondence are the first Laplace transforms of the intersector surfaces for the given correspondence.*

* Fubini, loc. cit., p. 249.

The corresponding statement can be made for the minus first Laplace transforms.

## 8. Reciprocal congruences

The Green reciprocal relation for a pair of congruences represents a particular type of the correspondence defined by (2.1). As a conclusion to the present discussion the equations of the correspondence between an arbitrary pair of reciprocal congruences may be exhibited.

Let the four homogeneous coördinates $x_k$ of an arbitrary point $P_x$ on a surface $S_x$ be given as analytic functions of two independent variables $u$ and $v$. Suppose further that $S_x$ is non-degenerate and non-developable, and is referred to its asymptotic net. Then the functions $x_k$ are a fundamental set of solutions of a completely integrable system of partial differential equations which may be reduced to the Fubini form

$$x_{uu} = px + \theta_u x_u + \beta x_v, \quad x_{vv} = qx + \gamma x_u + \theta_v x_v, \tag{8.1}$$

where $\theta = \log \beta\gamma$, and where certain integrability conditions are satisfied.

The points $P_\rho, P_\sigma$ defined by

$$\rho = x_u + bx, \quad \sigma = x_v + ax, \tag{8.2}$$

where $a$ and $b$ are functions of $u$ and $v$, are evidently points which lie in the tangent plane of $S_x$ at $P_x$. The reciprocal of $l_{\rho\sigma}$ in the osculating quadric of $S_x$ at $P_x$ is the line $l_{xy}$ which joins $P_x$ and $P_y$, the latter being defined by

$$y = ax_u + bx_v + x_{uv}. \tag{8.3}$$

Two congruences $\Gamma_{xy}, \Gamma_{\rho\sigma}$ have thus been determined in one-to-one correspondence, so that corresponding generators do not intersect. The coefficients $a^{ij}$ of the system (2.1) for the correspondence are as follows:

$$\begin{array}{llll} a^{11} = -b, & a^{12} = 0, & a^{13} = 1, & a^{14} = 0, \\ a^{21} = P, & a^{22} = \theta_u + b, & a^{23} = A, & a^{24} = F', \\ a^{31} = F, & a^{32} = 0, & a^{33} = \theta_u + b, & a^{34} = \beta, \\ a^{41} = a_u + ab, & a^{42} = 1, & a^{43} = 0, & a^{44} = -b, \end{array} \tag{8.4}$$

where we have placed

$$\begin{aligned} A &= \theta_{uv} + a_u - ab + \beta\gamma, \\ F &= p - b\theta_u - \beta a + b_u - b^2, \\ F' &= F + 2a\beta + \beta\theta_v, \\ P &= a(p - F') - Ab + p_v + \beta q. \end{aligned} \tag{8.5}$$

The coefficients $b^{ij}$ of (2.1) may be found from (8.4) and (8.5) by means of the substitution

$$\begin{pmatrix} a & \beta & u & 3 & A & F & F' & P \\ b & \gamma & v & 4 & B & G & G' & Q \end{pmatrix}. \tag{8.6}$$

The coefficients of (5.3) taken in conjunction with (8.4) and (8.6) are in substance those given by Lane* for his theory of pairs of reciprocal ruled surfaces.

* Lane, loc. cit., p. 287. Lane here uses the Wilczynski form of the equations for $S_2$.

UNIVERSITY OF ALBERTA,
EDMONTON, CANADA

# ON THE SOLUTION OF THE EQUATIONS OF ELASTIC EQUILIBRIUM SUITABLE FOR ELLIPTIC BOUNDARIES*

BY

SUDDHODAN GHOSH

1. The solution of the problem of the equilibrium of an isotropic elastic body with a given boundary depends, in the absence of body forces, on the solution of the partial differential equations

$$(\lambda + \mu)\frac{\partial \Delta}{\partial x} + \mu \nabla^2 u = 0,$$

$$(\lambda + \mu)\frac{\partial \Delta}{\partial y} + \mu \nabla^2 v = 0, \tag{1}$$

$$(\lambda + \mu)\frac{\partial \Delta}{\partial z} + \mu \nabla^2 w = 0,$$

where

$$\Delta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \tag{2}$$

subject to given conditions at the boundary. But the determination of the three functions $u$, $v$, $w$ satisfying equations (1) and (2) and the conditions at the bounding surface is a matter of great difficulty and hitherto such functions have been discovered only in the case of a few boundaries. It frequently happens, as with the kindred problems in hydrodynamics and electrostatics, that a problem which cannot be solved in three dimensions readily admits of a two-dimensional solution. Though these two-dimensional analogues can only be imperfectly produced in nature, they are of great importance on account of the light they throw on the more general problems which cannot be solved completely. From this point of view, the solutions of equations (1) and (2) in two dimensions are interesting and instructive and have actually received the attention of a large number of investigators.

It was shown by Airy that in the case of plane strain, the stresses can be expressed in terms of a stress function $\chi$ by the formulas

$$X_x = \frac{\partial^2 \chi}{\partial y^2}, \quad Y_y = \frac{\partial^2 \chi}{\partial x^2}, \quad X_y = -\frac{\partial^2 \chi}{\partial x \partial y}, \tag{3}$$

* Presented to the Society, February 25, 1928; received by the editors in January, 1928.

which in the absence of body forces further satisfies the partial differential equation of the fourth order

$$\nabla_1^4 \chi = 0 \tag{4}$$

where

$$\nabla_1^4 = \nabla_1^2 \cdot \nabla_1^2 \quad \text{and} \quad \nabla_1^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Consequently, equation (4) has received a considerable amount of attention from various mathematicians. The solutions in cartesian coördinates and their application to the problem of flexure of rectangular and cylindrical beams are well known. Those in polar coördinates have been given by Michell* and have been applied to plane bodies with concentric circular boundaries. Lately, a solution in bipolar coördinates has also been given † thus opening the way of the consideration of areas bounded by two non-concentric circles. In the present paper a solution in elliptic coördinates has been found and applied to two problems of plane strain. A single-valued solution is found suitable for the case of an elliptic cylindrical cavity in an infinite solid while another many-valued solution applicable to the state of strain in an elliptic cylindrical shell which has suffered dislocation due to a triangular axial fissure has also been found. Love‡ has given an outline of a method of solving problems in plane strain in elliptic coördinates, but his method is only applicable to cases where the surface displacements are given. The method developed in this paper lends itself easily to the construction of the solution of the particular problem dealt with by Love, as will be shown in §5. Some other applications of elliptic coördinates are given by Kolosoff,§ Pöschl,‖ and Inglis.¶

2. Let us use the transformation given by

$$x + iy = c \cosh (\alpha + i\beta), \tag{5}$$

so that

$$x = c \cosh \alpha \cos \beta, \quad y = c \sinh \alpha \sin \beta, \tag{6}$$

---

* Proceedings of the London Mathematical Society, vol. 31 (1900), p. 100.

† Jeffery, Philosophical Transactions, Royal Society, (A), vol. 221, pp. 265–293.

‡ *Theory of Elasticity* (3d edition), p. 275.

§ Zeitschrift für Mathematik und Physik, vol. 62 (1914), p. 384.

‖ Mathematische Zeitschrift, vol. 11 (1921), p. 89.

¶ Transactions of the Institute of Naval Architects, vol. 55, part 1, p. 219. My attention was drawn to Inglis's paper by one of the referees with the remark that it would be interesting to compare the expressions obtained in §4 of this paper with those obtained by Inglis; but unfortunately as the paper was not available here no such comparison could be made.

and

$$\frac{1}{h^2} = \left(\frac{\partial x}{\partial \alpha}\right)^2 + \left(\frac{\partial y}{\partial \alpha}\right)^2 = \frac{c^2}{2}(\cosh 2\alpha - \cos 2\beta).$$

The curves $\alpha =$ constant and $\beta =$ constant are confocal ellipses and hyperbolas respectively. The curve $\alpha = 0$ is the portion of the $x$-axis between the foci, while on the right-hand side of the $x$-axis between the focus and infinity, $\beta = 0$, and on the left-hand side between the other focus and infinity, $\beta = \pi$. On the positive side of the $y$-axis $\beta = \pi/2$ and on the negative side $\beta = 3\pi/2$.

We have therefore a set of coördinates suitable for the treatment of two-dimensional problems in which the region concerned is bounded by confocal ellipses and hyperbolas.

The stress components are given by equation (3). Transforming into curvilinear coördinates $(\alpha, \beta)$, the stress system can be written as

$$\widehat{\alpha\alpha} = h\frac{\partial}{\partial \beta}\left(h\frac{\partial \chi}{\partial \beta}\right) - h\frac{\partial h}{\partial \alpha}\frac{\partial \chi}{\partial \alpha},$$

$$\widehat{\beta\beta} = h\frac{\partial}{\partial \alpha}\left(h\frac{\partial \chi}{\partial \alpha}\right) - h\frac{\partial h}{\partial \beta}\frac{\partial \chi}{\partial \beta}, \tag{7}$$

$$\widehat{\alpha\beta} = -h\frac{\partial}{\partial \alpha}\left(h\frac{\partial \chi}{\partial \beta}\right) - h\frac{\partial h}{\partial \beta}\frac{\partial \chi}{\partial \alpha}.$$

In the present case we have

$$\frac{\widehat{\alpha\alpha}}{h^4} = \frac{1}{h^2}\frac{\partial^2 \chi}{\partial \beta^2} + \frac{1}{2}\frac{\partial}{\partial \alpha}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \beta}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \beta}$$

$$= \frac{c^2}{2}\left[(\cosh 2\alpha - \cos 2\beta)\frac{\partial^2 \chi}{\partial \beta^2} + \sinh 2\alpha \frac{\partial \chi}{\partial \alpha} - \sin 2\beta \frac{\partial \chi}{\partial \beta}\right];$$

$$\frac{\widehat{\beta\beta}}{h^4} = \frac{1}{h^2}\frac{\partial^2 \chi}{\partial \alpha^2} - \frac{1}{2}\frac{\partial}{\partial \alpha}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \alpha} + \frac{1}{2}\frac{\partial}{\partial \beta}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \beta} \tag{8}$$

$$= \frac{c^2}{2}\left[(\cosh 2\alpha - \cos 2\beta)\frac{\partial^2 \chi}{\partial \alpha^2} - \sinh 2\alpha \frac{\partial \chi}{\partial \alpha} + \sin 2\beta \frac{\partial \chi}{\partial \beta}\right];$$

$$\frac{\widehat{\alpha\beta}}{h^4} = -\frac{1}{h^2}\frac{\partial^2 \chi}{\partial \alpha \partial \beta} + \frac{1}{2}\frac{\partial}{\partial \alpha}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \beta} + \frac{1}{2}\frac{\partial}{\partial \beta}\left(\frac{1}{h^2}\right)\frac{\partial \chi}{\partial \alpha}$$

$$= \frac{c^2}{2}\left[-(\cosh 2\alpha - \cos 2\beta)\frac{\partial^2 \chi}{\partial \alpha \partial \beta} + \sinh 2\alpha \frac{\partial \chi}{\partial \beta} + \sin 2\beta \frac{\partial \chi}{\partial \alpha}\right].$$

The displacements are given by*

* Jeffery, loc. cit.

$$2\mu\frac{u}{h} = -\frac{\partial\chi}{\partial\alpha}+\frac{1}{h^2}\frac{\partial P}{\partial\beta}, \quad 2\mu\frac{v}{h} = -\frac{\partial\chi}{\partial\beta}+\frac{1}{h^2}\frac{\partial P}{\partial\alpha}, \tag{9}$$

where $P$ satisfies the equations

$$\nabla_1^2 P = 0 \tag{10}$$

and

$$\frac{\partial}{\partial\alpha}\left(\frac{1}{h^2}\frac{\partial P}{\partial\beta}\right)+\frac{\partial}{\partial\beta}\left(\frac{1}{h^2}\frac{\partial P}{\partial\alpha}\right) = \frac{\lambda+2\mu}{\lambda+\mu}\left(\frac{\partial^2\chi}{\partial\alpha^2}+\frac{\partial^2\chi}{\partial\beta^2}\right). \tag{11}$$

The stress function noted above satisfies equation (4).

3. Let us now proceed with the solution of equation (4). Writing it in the form

$$h^2\left(\frac{\partial^2}{\partial\alpha^2}+\frac{\partial^2}{\partial\beta^2}\right)h^2\left(\frac{\partial^2\chi}{\partial\alpha^2}+\frac{\partial^2\chi}{\partial\beta^2}\right) = 0,$$

we notice that

$$h^2\left(\frac{\partial^2\chi}{\partial\alpha^2}+\frac{\partial^2\chi}{\partial\beta^2}\right)$$

is a plane harmonic function.

We now take the following form for the harmonic:

$$h^2\left(\frac{\partial^2\chi}{\partial\alpha^2}+\frac{\partial^2\chi}{\partial\beta^2}\right) = \frac{2}{c^2}\Big[A_0 + A\alpha + B\beta + 2\sum_{n=1}^{n=\infty}(A_n\cosh n\alpha + B_n\sinh n\alpha)\cos n\beta + 2\sum_{n=1}^{n=\infty}(C_n\cosh n\alpha + D_n\sinh n\alpha)\sin n\beta\Big]. \tag{12}$$

Hence

$$\begin{aligned}\frac{\partial^2\chi}{\partial\alpha^2}+\frac{\partial^2\chi}{\partial\beta^2} &= (A\alpha+B\beta)(\cosh 2\alpha-\cos 2\beta)+(A_0-A_2)\cosh 2\alpha - B_2\sinh 2\alpha\\ &+\{-2B_1\sinh\alpha+(A_1-A_3)\cosh 3\alpha+(B_1-B_3)\sinh 3\alpha\}\cos\beta\\ &+\{2C_1\cosh\alpha+(C_1-C_3)\cosh 3\alpha+(D_1-D_3)\sinh 3\alpha\}\sin\beta\\ &+\{-(A_0-A_2)+(A_2-A_4)\cosh 4\alpha+(B_2-B_4)\sinh 4\alpha\}\cos 2\beta\\ &+\{C_2+(C_2-C_4)\cosh 4\alpha+(D_2-D_4)\sinh 4\alpha\}\sin 2\beta\\ &+\sum_{n=3}^{n=\infty}[-(A_{n-2}-A_n)\cosh(n-2)\alpha-(B_{n-2}-B_n)\sinh(n-2)\alpha\\ &+(A_n-A_{n+2})\cosh(n+2)\alpha+(B_n-B_{n+2})\sinh(n+2)\alpha]\cos n\beta\\ &+\sum_{n=3}^{n=\infty}[-(C_{n-2}-C_n)\cosh(n-2)\alpha-(D_{n-2}-D_n)\sinh(n-2)\alpha\\ &+(C_n-C_{n+2})\cosh(n+2)\alpha+(D_n-D_{n+2})\sinh(n+2)\alpha]\sin n\beta.\end{aligned} \tag{13}$$

We try to satisfy this equation by assuming the following form for $\chi$:

$$\chi = \frac{A}{4}[\alpha(\cosh 2\alpha + \cos 2\beta) - \sinh 2\alpha] + \frac{B}{4}[\beta(\cosh 2\alpha + \cos 2\beta) - \sin 2\beta]$$

$$+ \phi_0 + \sum_{n=1}^{n=\infty} \phi_n \cos n\beta + \sum_{n=1}^{n=\infty} \psi_n \sin n\beta + \chi', \tag{14}$$

where $\phi$'s and $\psi$'s are functions of $\alpha$ only and $\chi'$ is a plane harmonic such as

$$A'\alpha + B'\beta + \sum_{n=1}^{n=\infty} (A_n' \cosh n\alpha + B_n' \sinh n\alpha) \cos n\beta$$

$$+ \sum_{n=1}^{n=\infty} (C_n' \cosh n\alpha + D_n' \sinh n\alpha) \sin n\beta.$$

Substituting in equation (13) and comparing coefficients we get the values of $\phi$'s and $\psi$'s:

$$\phi_0 = \frac{1}{4}(A_0 - A_2) \cosh 2\alpha - \frac{1}{4}B_2 \sinh 2\alpha,$$

$$\phi_1 = \frac{1}{8}[(A_1 - A_3) \cosh 3\alpha + (B_1 - B_3) \sinh 3\alpha] - B_1\alpha \cosh \alpha,$$

$$\phi_2 = \frac{1}{4}(A_0 - A_2) + \frac{1}{12}[(A_2 - A_4) \cosh 4\alpha + (B_2 - B_4) \sinh 4\alpha],$$

$$\phi_n = \frac{1}{4(n-1)}[(A_{n-2} - A_n) \cosh (n-2)\alpha + (B_{n-2} - B_n) \sinh (n-2)\alpha]$$

$$+ \frac{1}{4(n+1)}[(A_n - A_{n+2}) \cosh (n+2)\alpha + (B_n - B_{n+2}) \sinh (n+2)\alpha];$$

$$\psi_1 = \frac{1}{8}[(C_1 - C_3) \cosh 3\alpha + (D_1 - D_3) \sinh 3\alpha] + C_1\alpha \sinh \alpha,$$

$$\psi_2 = -\frac{1}{4}C_2 + \frac{1}{12}[(C_2 - C_4) \cosh 4\alpha + (D_2 - D_4) \sinh 4\alpha],$$

$$\psi_n = \frac{1}{4(n-1)}[(C_{n-2} - C_n) \cosh (n-2)\alpha + (D_{n-2} - D_n) \sinh (n-2)\alpha]$$

$$+ \frac{1}{4(n+1)}[(C_n - C_{n+2}) \cosh (n+2)\alpha + (D_n - D_{n+2}) \sinh (n+2)\alpha].$$

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

It is evident that the terms $\cosh \alpha \cos \beta$ and $\sinh \alpha \sin \beta$ in $\chi'$ represent a displacement that is possible in a rigid body.

Hence omitting the rigid body displacement terms and also the terms which ultimately make the stresses many-valued, we can write

$$\chi = A\alpha(\cosh 2\alpha + \cos 2\beta) + B\beta + \phi_0 + \sum_{n=1}^{n=\infty} \phi_n \cos n\beta + \sum_{n=1}^{n=\infty} \psi_n \sin n\beta, \tag{15}$$

where

$$\begin{aligned}
\phi_0 &= a_0 \cosh 2\alpha + b \sinh 2\alpha + c\alpha, \\
\phi_1 &= a'\alpha \cosh \alpha + a_1 \cosh 3\alpha + b_1 \sinh 3\alpha + b_1' \sinh \alpha, \\
\phi_2 &= a_0 + a_2 \cosh 4\alpha + b_2 \sinh 4\alpha + a_2' \cosh 2\alpha + b_2' \sinh 2\alpha, \\
\phi_n &= a_{n-2} \cosh (n-2)\alpha + b_{n-2} \sinh (n-2)\alpha + a_n \cosh (n+2)\alpha \\
&\quad + b_n \sinh (n+2)\alpha + a_n' \cosh n\alpha + b_n' \sinh n\alpha,
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\psi_1 &= c'\alpha \sinh \alpha + c_1 \cosh 3\alpha + d_1 \sinh 3\alpha + c_1' \cosh \alpha, \\
\psi_2 &= c_0 + c_2 \cosh 4\alpha + d_2 \sinh 4\alpha + c_2' \cosh 2\alpha + d_2' \sinh 2\alpha, \\
\psi_n &= c_{n-2} \cosh (n-2)\alpha + d_{n-2} \sinh (n-2)\alpha + c_n \cosh (n+2)\alpha \\
&\quad + d_n \sinh (n+2)\alpha + c_n' \cosh n\alpha + d_n' \sinh n\alpha.
\end{aligned} \tag{17}$$

4. Even when $\chi$ is known the displacements $u$ and $v$ can only be obtained by first determining $P$ which satisfies equations (10) and (11). Now, instead of finding $P$ directly, it will be convenient to calculate

$$\frac{1}{h^2}\frac{\partial P}{\partial \alpha} \quad \text{and} \quad \frac{1}{h^2}\frac{\partial P}{\partial \beta}$$

subject to the conditions (10) and (11).

It will subsequently be shown that the terms

$$\alpha(\cosh 2\alpha + \cos 2\beta), \quad \alpha \cosh \alpha \cos \beta \quad \text{and} \quad \alpha \sinh \alpha \sin \beta$$

give rise to many-valued *displacements*, so that omitting them from $\chi$ and taking into account the fact that $P$ is a harmonic function we can write

$$\begin{aligned}\frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= A_0 \sinh 2\alpha + D_0 \cosh 2\alpha + (B' \cosh \alpha + A_1 \sinh 3\alpha \\ &+ B_1 \cosh 3\alpha) \sin \beta + (C \sinh \alpha + C_1 \sinh 3\alpha + D_1 \cosh 3\alpha) \cos \beta \\ &+ (B_0 + A_2 \sinh 4\alpha + B_2 \cosh 4\alpha) \sin 2\beta + (- D_0 + C_2 \sinh 4\alpha \\ &+ D_2 \cosh 4\alpha) \cos 2\beta + \sum_{n=3}^{n=\infty} [\{- A_{n-2} \sinh (n-2)\alpha \\ &- B_{n-2} \cosh (n-2)\alpha + A_n \sinh (n+2)\alpha + B_n \cosh (n+2)\alpha\} \sin n\beta \\ &+ \{- C_{n-2} \sinh (n-2)\alpha - D_{n-2} \cosh (n-2)\alpha + C_n \sinh (n+2)\alpha \\ &+ D_n \cosh (n+2)\alpha\} \cos n\beta],\end{aligned} \tag{18}$$

and

$$\begin{aligned}\frac{1}{h^2}\frac{\partial P}{\partial \beta} &= - B_0 \sinh 2\alpha + C_0 \cosh 2\alpha + (- B' \sinh \alpha + A_1 \cosh 3\alpha \\ &+ B_1 \sinh 3\alpha) \cos \beta + (C \cosh \alpha - C_1 \cosh 3\alpha - D_1 \sinh 3\alpha) \sin \beta \\ &+ (- C_0 + A_2 \cosh 4\alpha + B_2 \sinh 4\alpha) \cos 2\beta + (A_0 - C_2 \cosh 4\alpha \\ &- D_2 \sinh 4\alpha) \sin 2\beta + \sum_{n=3}^{n=\infty} [\{- A_{n-2} \cosh (n-2)\alpha \\ &- B_{n-2} \sinh (n-2)\alpha + A_n \cosh (n+2)\alpha \\ &+ B_n \sinh (n+2)\alpha\} \cos n\beta + \{C_{n-2} \cosh (n-2)\alpha \\ &+ D_{n-2} \sinh (n-2)\alpha - C_n \cosh (n+2)\alpha - D_n \sinh (n+2)\alpha\} \sin n\beta].\end{aligned} \tag{19}$$

Here the general plane harmonic function $P$ has been differentiated with respect to $\alpha$ and $\beta$ and multiplied by $1/h^2$.

Now substituting in (11) and equating coefficients as usual, we find the values of the coefficients $A_n, B_n, C_n, D_n, \cdots$, as follows:

$$\begin{aligned} &B_0 = - 2ka_0, \quad C_0 = 2kb, \quad D_0 = - 2kc_0, \\ &A_n = 2kb_n, \quad B_n = 2ka_n, \quad C_n = - 2kd_n, \quad D_n = - 2kc_n, \end{aligned}$$

where

$$k = \frac{\lambda + 2\mu}{\lambda + \mu}.$$

In this way all the constants $A_n, B_n, C_n, D_n$ can be determined except $A_0, B'$ and $C$. The terms containing $A_0, B'$ and $C$ give rise to a displacement which has the following components parallel to axes of $x$ and $y$ respectively:

$$\frac{1}{2\mu c^2}(- B'c - 2A_0 y) \quad \text{and} \quad \frac{1}{2\mu c_2}(Cc + 2A_0 x),$$

thus showing that the displacement is one possible in a rigid body. So we can omit $A_0$, $B'$ and $C$. There still remain terms in $\chi$, e.g. those involving $a'$, $c'$ and $A$ which are not represented in the differential coefficients $(1/h^2)(\partial P/\partial\alpha)$ and $(1/h^2)(\partial P/\partial\beta)$. It is necessary to consider them separately. For the term $\alpha \cosh \alpha \cos \beta$ in $\chi$, let us assume

$$\begin{aligned} \frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= A'(\beta \sinh \alpha \cos \beta - \alpha \cosh \alpha \sin \beta), \\ \frac{1}{h^2}\frac{\partial P}{\partial \beta} &= A'(\beta \cosh \alpha \sin \beta + \alpha \sinh \alpha \cos \beta). \end{aligned} \tag{20}$$

It can easily be shown that these satisfy

$$\nabla_1^2 P = 0$$

and that

$$\frac{\partial}{\partial \alpha}\left(\frac{1}{h^2}\frac{\partial P}{\partial \beta}\right) + \frac{\partial}{\partial \beta}\left(\frac{1}{h^2}\frac{\partial P}{\partial \alpha}\right) = 2A' \sinh \alpha \cos \beta.$$

$A'$ can thus be determined.

Similarly for the other term $\alpha \sinh \alpha \sin \beta$, we can take

$$\begin{aligned} \frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= B'(\alpha \sinh \alpha \cos \beta + \beta \cosh \alpha \sin \beta), \\ \frac{1}{h^2}\frac{\partial P}{\partial \beta} &= B'(\alpha \cosh \alpha \sin \beta - \beta \sinh \alpha \cos \beta). \end{aligned} \tag{21}$$

For the term $\alpha(\cosh 2\alpha + \cos 2\beta)$, we take

$$\begin{aligned} \frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= c'(\beta \sinh 2\alpha - \alpha \sin 2\beta), \\ \frac{1}{h^2}\frac{\partial P}{\partial \beta} &= c'(\beta \sin 2\beta + \alpha \sinh 2\alpha). \end{aligned} \tag{22}$$

They satisfy equation (10) and give

$$\frac{\partial}{\partial \alpha}\left(\frac{1}{h^2}\frac{\partial P}{\partial \beta}\right) + \frac{\partial}{\partial \beta}\left(\frac{1}{h^2}\frac{\partial P}{\partial \alpha}\right) = 2c' \sinh 2\alpha + 2c'\alpha(\cosh 2\alpha - \cos 2\beta).$$

From (11) it is easily seen that

$$c' = \frac{2(\lambda + 2\mu)}{\lambda + \mu} A.$$

The stresses can be easily calculated from (8) as below:

$$
\begin{aligned}
\frac{2}{c^2}\frac{\widehat{\alpha\alpha}}{h^4} &= (3\phi_2 + \phi_0' \sinh 2\alpha) + (-\phi_1 \cosh 2\alpha + \phi_1' \sinh 2\alpha + 6\phi_3)\cos\beta \\
&\quad + (-\psi_1 \cosh 2\alpha + \psi_1' \sinh 2\alpha + 6\psi_3)\sin\beta - B \sin 2\beta \\
&\quad + \sum_{n=2}^{n=\infty}\Big[\Big\{-n^2\phi\cosh 2\alpha + \phi_n' \sinh 2\alpha + \frac{1}{2}(n+2)(n+3)\phi_{n+2} \\
&\quad + \frac{1}{2}(n-2)(n-3)\phi_{n-2}\Big\}\cos n\beta + \Big\{n^2\psi_n \cosh 2\alpha + \psi_n' \sinh 2\alpha \\
&\quad + \frac{1}{2}(n+2)(n+3)\psi_{n+2} + \frac{1}{2}(n-2)(n-3)\psi_{n-2}\Big\}\sin n\beta\Big];
\end{aligned} \tag{23}
$$

$$
\begin{aligned}
\frac{2}{c^2}\frac{\widehat{\beta\beta}}{h^4} &= \phi_0'' \cosh 2\alpha - \phi_0' \sinh 2\alpha - \frac{1}{2}\phi_2' - \phi_2 \\
&\quad + \Big(\phi_1'' \cosh 2\alpha - \phi_1' \sinh 2\alpha - \frac{1}{2}\phi_3'' - \frac{3}{2}\phi_3\Big)\cos\beta \\
&\quad + \Big(\psi_1'' \cosh 2\alpha - \psi_1' \sinh 2\alpha - \frac{1}{2}\psi_3'' - \frac{3}{2}\psi_3\Big)\sin\beta + B\sin 2\beta \\
&\quad + \sum_{n=2}^{n=\infty}\Big[\Big\{-\frac{1}{2}\phi_{n-2}'' + \frac{1}{2}(n-2)\phi_{n-2} + \phi_n'' \cosh 2\alpha - \phi_n' \sinh 2\alpha \\
&\quad - \frac{1}{2}\phi_{n+2}'' - \frac{1}{2}(n+2)\phi_{n+2}\Big\}\cos n\beta + \Big\{-\frac{1}{2}\psi_{n-2}'' + \frac{1}{2}(n-2)\psi_{n-2} \\
&\quad + \psi_n'' \cosh 2\alpha - \psi_n' \sinh 2\alpha - \frac{1}{2}\psi_{n+2}'' - \frac{1}{2}(n+2)\psi_{n+2}\Big\}\sin n\beta\Big];
\end{aligned} \tag{24}
$$

and

$$
\begin{aligned}
\frac{2}{c^2}\frac{\widehat{\alpha\beta}}{h^4} &= B\sinh 2\alpha + (\phi_1' \cosh 2\alpha - \phi_1 \sinh 2\alpha - 2\phi_3')\sin\beta + (-\psi_1' \cosh 2\alpha \\
&\quad + \psi_1 \sinh 2\alpha + 2\psi_3')\cos\beta + \Big(\phi_0' + 2\phi_2' \cosh 2\alpha - 2\phi_2 \sinh 2\alpha \\
&\quad - \frac{5}{2}\phi_4'\Big)\sin 2\beta + \Big(-2\psi_2' \cosh 2\alpha + 2\psi_2 \sinh 2\alpha + \frac{5}{2}\psi_4'\Big)\cos 2\beta \\
&\quad + \sum_{n=3}^{n=\infty}\Big[\Big\{-\frac{1}{2}(n-3)\phi_{n-2}' + n\phi_n' \cosh 2\alpha - n\phi_n \sinh 2\alpha \\
&\quad - \frac{1}{2}(n+3)\phi_{n+2}'\Big\}\sin n\beta + \Big\{\frac{1}{2}(n-3)\psi_{n-2}' - n\psi_n' \cosh 2\alpha \\
&\quad + n\psi_n \sinh 2\alpha + \frac{1}{2}(n+3)\psi_{n+2}'\Big\}\cos n\beta\Big].
\end{aligned} \tag{25}
$$

Here we have omitted the first term in $\chi$ given in (15) while calculating the stresses.

5. We shall now consider some applications of the above solutions.

(a). Let an infinite solid with an elliptic cylindrical hollow be subjected to a pressure $P_1$ on the inner elliptic boundary.*

Assume in this case

$$\chi = A(e^{-2\alpha} + \cos 2\beta) + B\alpha. \tag{26}$$

Then we have

$$\frac{c^2}{2}(\cosh 2\alpha - \cos 2\beta)^2\widehat{\alpha\alpha} = \{A(3 - 2e^{-2\alpha}\sinh 2\alpha) + B\sinh 2\alpha\} - 4A\cosh 2\alpha\cos 2\beta + A\cos 4\beta,$$

and

$$\frac{c^2}{2}(\cosh 2\alpha - \cos 2\beta)^2\widehat{\alpha\beta} = (B - 2A\cosh 2\alpha)\sin 2\beta.$$

The conditions on the boundary give

$$\widehat{\alpha\alpha} = -P_1 \quad \text{and} \quad \widehat{\alpha\beta} = 0 \quad \text{when} \quad \alpha = \alpha_1,$$

also

$$\widehat{\alpha\alpha} = 0 \quad \text{and} \quad \widehat{\alpha\beta} = 0 \quad \text{when} \quad \alpha = \infty.$$

These give

$$A = -\frac{P_1c^2}{4} \quad \text{and} \quad B = -\frac{P_1c^2}{2}\cosh 2\alpha_1. \tag{27}$$

In this case $P$ can be easily shown to be equal to

$$-\frac{\lambda + 2\mu}{\lambda + \mu}P_1\tan^{-1}\left(\frac{e^{-2\alpha}\sin 2\beta}{1 - e^{-2\alpha}\cos 2\beta}\right).$$

Equations (9) now give the following displacements:

$$\frac{u}{h} = \frac{P_1c^2}{4(\lambda + \mu)}e^{-2\alpha} - \frac{\lambda + 2\mu}{4\mu(\lambda + \mu)}P_1c^2\cos 2\beta + \frac{P_1c^2}{4\mu}\cosh 2\alpha_1,$$
$$\frac{v}{h} = \frac{P_1c^2}{4(\lambda + \mu)}\sin 2\beta. \tag{28}$$

(b). For the second case we are going to consider let us assume

* This case differs from the cases studied by Kolosoff and Pöschl in boundary conditions only. In Pöschl's case $\chi$=const. and $\partial\chi/\partial\alpha$=0 on the inner boundary $\alpha=\alpha_1$ and $\chi\to(1/16)pc^2e^{2\alpha}[1+\cos 2(\beta-\beta')]$ as $\alpha\to\infty$.

$$\chi = (A + A'\alpha)(\cosh 2\alpha + \cos 2\beta) + B \sinh 2\alpha + c\alpha + (D \cosh 2\alpha + E \sinh 2\alpha) \cos 2\beta. \tag{29}$$

If the boundaries $\alpha=\alpha_1$ and $\alpha=\alpha_2$ be subjected to pressures $P_1$ and $P_2$ respectively, we easily find by calculating the stresses and taking into account the boundary conditions the values

$$\begin{aligned}
A &= -\frac{1}{4M}c^2\left[P_1\left\{\alpha_2 - \frac{1}{2}\tanh(\alpha_2 - \alpha_1)\right\} - P_2\left\{\alpha_1 - \frac{1}{2}\tanh(\alpha_1 - \alpha_2)\right\}\right],\\
A' &= \frac{1}{4M}(P_1 - P_2)c^2,\\
B &= -\frac{1}{4M}(P_1 - P_2)c^2,\\
C &= \frac{1}{4M}(P_1 - P_2)c^2\frac{\cosh(\alpha_2 + \alpha_1)}{\cosh(\alpha_2 - \alpha_1)},\\
D &= \frac{1}{8M}(P_1 - P_2)c^2\frac{\sinh(\alpha_2 + \alpha_1)}{\cosh(\alpha_2 - \alpha_1)},\\
E &= -\frac{1}{8M}(P_1 - P_2)c^2\frac{\cosh(\alpha_2 + \alpha_1)}{\cosh(\alpha_2 - \alpha_1)},
\end{aligned} \tag{30}$$

where

$$M = (\alpha_2 - \alpha_1) - \tanh(\alpha_2 - \alpha_1).$$

Moreover, we have in this case

$$\begin{aligned}
\frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= \frac{2(\lambda + 2\mu)}{\lambda + \mu}[A'(\beta \sinh 2\alpha - \alpha \sin 2\beta) - A \sin 2\beta],\\
\frac{1}{h^2}\frac{\partial P}{\partial \beta} &= \frac{2(\lambda + 2\mu)}{\lambda + \mu}[A'(\beta \sin 2\beta + \alpha \sinh 2\alpha) + A \sinh 2\alpha\\
&\qquad + B(\cosh 2\alpha - \cos 2\beta)].
\end{aligned}$$

Also the displacements are given by

$$\begin{aligned}
\frac{u}{h} = \frac{1}{\lambda + \mu}\Bigg[&A \sinh 2\alpha + \left(B - \frac{\lambda + \mu}{\mu}A'\right)\cosh 2\alpha + \frac{\lambda + 2\mu}{\mu}A'\beta \sin 2\beta\\
&+ A'\alpha \sinh 2\alpha - \left(\frac{\lambda + 2\mu}{\mu}B + \frac{\lambda + \mu}{2\mu}A'\right)\cos 2\beta\Bigg] - \frac{C}{2\mu}\\
&- \frac{1}{\mu}(D \sinh 2\alpha + E \cosh 2\alpha)\cos 2\beta,
\end{aligned} \tag{31}$$

and

$$\frac{v}{h} = \frac{1}{\lambda+\mu}\left[\frac{\lambda+2\mu}{\mu}A'\beta \sinh 2\alpha - A'\alpha \sin 2\beta - A \sin 2\beta\right] + \frac{1}{\mu}(D \cosh 2\alpha + E \sinh 2\alpha) \sin 2\beta. \tag{32}$$

These displacements are not single-valued. Considering only the many-valued terms, we write

$$\frac{u'}{h} = \frac{\lambda+2\mu}{\mu(\lambda+\mu)}A'\beta \sin 2\beta, \qquad \frac{v'}{h} = \frac{\lambda+2\mu}{\mu(\lambda+\mu)}A'\beta \sinh 2\alpha. \tag{33}$$

If we restrict $\beta$ by a convention such as $0<\beta<2\pi$, these displacements become single-valued and discontinuous. When $\beta=0$, we have $u'=0$ and $v'=0$; when $\beta=2\pi$, we have

$$u' = 0 \quad \text{and} \quad v' = \frac{4\pi(\lambda+2\mu)}{\mu(\lambda+\mu)}\frac{A'}{c}\cosh\alpha.$$

Therefore there is no discontinuity in the value of $u$, but the value of $v$ decreases suddenly by the amount $v_0$ given by

$$v_0 = \frac{4\pi(\lambda+2\mu)}{\mu(\lambda+\mu)}\frac{A'}{c}\cosh\alpha \tag{34}$$

as $\beta$ increases from $2\pi-0$ to $2\pi+0$. This is the jump in the value of $v$ on the positive $x$-axis.

Now when $\beta=0$,

$$x = x_0 = c\cosh\alpha.$$

Therefore

$$\frac{v_0}{x_0} = \frac{4\pi(\lambda+2\mu)}{\mu(\lambda+\mu)}\frac{A'}{c^2} = \text{constant}. \tag{35}$$

This result we can interpret as a dislocation* due to an axial wedge-shaped fissure bounded by the planes

$$y = \pm\frac{2\pi(\lambda+2\mu)}{\mu(\lambda+\mu)}\frac{A'}{c^2}x, \tag{36}$$

where $A'$ is given by (30).

* For the theory of dislocation, see Volterra's paper *Sur l'équilibre des corps élastiques multiplement connexes*, Annales de l'Ecole Normale Supérieure, (3), vol. 24 (1907), pp. 401–517. See also Love's *Theory of Elasticity* (3d edition), p. 219.

Hence the assumption made in (29) solves the problem of dislocation for a wedge-shaped fissure in a body bounded by two confocal elliptic cylinders.

(c). As a third application let us consider the problem dealt with by Love.*

Let an infinite solid be bounded internally by an elliptic boundary of semiaxes $a$ and $b$ and let the inner boundary be turned through an angle $\phi$.

Let $\alpha=\alpha_1$ be the inner boundary. Then the displacement at any point on the boundary is $r\phi$, perpendicular to the radius vector drawn from the center of the elliptic section to the point, where $r$ = distance of the point from the center. This displacement has a component $-c\phi \sinh \alpha \sin \beta$ parallel to $Ox$ and $c\phi \cosh \alpha \cos \beta$ parallel to $Oy$. If $u$, $v$ be the displacements along and perpendicular to the normal to $\alpha$ = constant we have when $\alpha=\alpha_1$

$$\begin{aligned} \frac{u}{h} &= \frac{1}{2}c^2\phi \sin 2\beta, \\ \frac{v}{h} &= \frac{1}{2}c^2\phi \sinh 2\alpha_1. \end{aligned} \tag{37}$$

Assume

$$\chi = A\beta + B \sin 2\beta + Ce^{-2\alpha} \sin 2\beta. \tag{38}$$

Then

$$\begin{aligned} \frac{1}{h^2}\frac{\partial P}{\partial \alpha} &= A_0 \sinh 2\alpha - \frac{2(\lambda+2\mu)}{\lambda+\mu}B(\cosh 2\alpha - \cos 2\beta), \\ \frac{1}{h^2}\frac{\partial P}{\partial \beta} &= A_0 \sin 2\beta, \end{aligned} \tag{39}$$

and

$$\begin{aligned} 2\mu\frac{u}{h} &= 2Ce^{-2\alpha} \sin 2\beta + A_0 \sin 2\beta, \\ 2\mu\frac{v}{h} &= -A + A_0 \sinh 2\alpha - \frac{2(\lambda+2\mu)}{\lambda+\mu}B \cosh 2\alpha \\ &\quad + \frac{2\mu}{\lambda+\mu}B \cos 2\beta - 2Ce^{-2\alpha} \cos 2\beta. \end{aligned} \tag{40}$$

From the boundary conditions (37) and from the condition that $u$ and $v$ vanish at infinity we find

* *Theory of Elasticity* (3d edition), p. 275.

$$
\begin{aligned}
A &= -\frac{\mu(\lambda+2\mu)}{\lambda+3\mu}c^2\phi e^{-2\alpha_1} - \mu c^2\phi \sinh 2\alpha_1,\\
B &= \frac{1}{2}\,\frac{\mu(\lambda+\mu)}{\lambda+3\mu}c^2\phi,\\
C &= \frac{1}{2}\,\frac{\mu^2}{\lambda+3\mu}c^2\phi e^{2\alpha_1},\\
A_0 &= \frac{\mu(\lambda+2\mu)}{\lambda+3\mu}c^2\phi.
\end{aligned}
\tag{41}
$$

Then

$$
\begin{aligned}
\frac{u}{h} &= \frac{1}{2}\,\frac{\mu}{\lambda+3\mu}c^2\phi e^{2\alpha_1}\left[e^{-2\alpha}+\frac{\lambda+2\mu}{\mu}e^{-2\alpha_1}\right]\sin 2\beta,\\
\frac{v}{h} &= \frac{1}{2}c^2\phi \sinh 2\alpha_1 + \frac{1}{2}c^2\phi e^{2\alpha_1}(e^{-2\alpha_1}-e^{-2\alpha})\left\{\frac{\lambda+2\mu}{\lambda+3\mu}e^{-2\alpha_1}+\frac{\mu}{\lambda+3\mu}\cos 2\beta\right\}.
\end{aligned}
\tag{42}
$$

But $a=c\cosh\alpha_1$ and $b=c\sinh\alpha_1$, and consequently

$$
\begin{aligned}
\frac{u}{h} &= \frac{1}{2}\,\frac{\mu}{\lambda+3\mu}(a+b)^2\phi\left\{e^{-2\alpha}+\frac{\lambda+2\mu}{\mu}\,\frac{a-b}{a+b}\right\}\sin 2\beta,\\
\frac{v}{h} &= ab\phi+\frac{1}{2}(a+b)^2\phi\left(\frac{a-b}{a+b}-e^{-2\alpha}\right)\left\{\frac{\lambda+2\mu}{\lambda+3\mu}\,\frac{a-b}{a+b}\right.\\
&\qquad\left.+\frac{\mu}{\lambda+3\mu}\cos 2\beta\right\}.
\end{aligned}
\tag{43}
$$

This is the same result as found by Love.

In conclusion I wish to express my thanks to Dr. N. R. Sen for guidance and help in course of the work.

UNIVERSITY COLLEGE OF SCIENCE,
CALCUTTA, INDIA

# ON THE UNILATERAL EQUATION IN MATRICES*

BY

WILLIAM E. ROTH

The equation

$$\sum_{i=0}^{p} A_i X^{p-i} = 0, \tag{1}$$

in which the coefficients $A_i$, $i=0, 1, 2, \cdots, p$, are $m \times n$ matrices having $m$ rows and $n$ columns of scalar elements, and where $X$ is an $n \times n$ matrix, is said to be unilateral† in that the known matrices $A_i$, $i=0, 1, 2, \cdots, p$, occur as factors on the left (or on the right) of each term of its left member. In case $m=n$ the equation was studied by Sylvester in several papers of a series that appeared largely in the Johns Hopkins University Circulars of 1883 and 1884, the Philosophical Magazine, and the Comptes Rendus de l'Académie de Sciences.‡ The solution of particular types of equations in matrices that may be regarded as special cases under the above was further studied by Buchheim, Frobenius, Taber, Kreis, Cecioni, Wedderburn and others.§ The restrictions adopted heretofore were such that the given

* Presented to the Society, August 29, 1929; received by the editors in August, 1929.

† Sylvester, *Equations in matrices*, Johns Hopkins University Circulars, vol. 3 (1884), p. 122, or Mathematical Papers, vol. IV, pp. 152–153.

‡ See particularly *On the three laws of motion in the world of universal algebra*, Johns Hopkins University Circulars, vol. 3 (1884), pp. 33, 34, 57, or Mathematical Papers, vol. IV, pp. 146–151; *Sur l'équation quadratique de Hamilton en quaternions ou en matrices du second ordre*, Comptes Rendus, vol. 99 (1884), pp. 555–558, pp. 621–631, or Mathematical Papers, vol. IV, pp. 190–198. This paper will be referred to as E.Q.

§ Buchheim, *An extension of a theorem of Professor Sylvester's relating to matrices*, Philosophical Magazine, vol. 22 (1886), p. 173.

Frobenius, *Über lineare Substitutionen und bilineare Formen*, Crelle's Journal, vol. 84 (1878), pp. 1–63; we shall refer to this paper as L.S. Also, *Über die cogredienten Transformation der bilinearen Formen*, Sitzungsbericht der königlichen Preussischen Akademie der Wissenschaften, 1896. This paper gives solutions of $X^2=A$ expressible as polynomials in $A$.

Taber, *On the automorphic linear transformation of an alternate bilinear form*, Mathematische Annalen, vol. 46 (1895), pp. 561–583.

Kreis, *Contribution à la Théorie des Systèmes Linéaires*, Zurich Thesis, 1906, and *Auflösung der Gleichung* $X^m=A$, Vierteljahrschrift der Naturforschenden Gesellschaft in Zurich, vol. 53 (1908), pp. 366–376.

Cecioni, *Sopra alcune operazioni algebriche sulle matrici*, Annali della Reale Scuola Normale Superiore di Pisa, vol. 11 (1909), pp. 1–140.

Wedderburn, *The automorphic transformation of a bilinear form*, Annals of Mathematics, (2), vol. 23 (1923), pp. 122–134.

Roth, *A solution of the matric equation* $P(X)=A$, these Transactions, vol. 30 (1928), pp. 579–596.

Franklin, *Algebraic matric equations*. This paper has not appeared in print. An abstract appears in the Bulletin of the American Mathematical Society, vol. 35 (1929), p. 7.

equation might have at best matric coefficients every two of which are commutative. The only cases in which equation (1) has received attention as far as the writer is aware is in the algebra of quaternions and that given it by Sylvester.*

In the following pages we shall seek necessary conditions for the existence of a solution of (1) and at the same time we shall indicate how a solution, when it exists, may be computed. The literature gives very essential aids in the attack of this problem when solutions whose elements belong to a restricted field of rationality are sought, and we have the work of Cecioni† as an example of such an attack upon a related problem; we shall nevertheless consider only solutions, $X$, of the given equation whose elements belong to the complex number field of rationality. The results we derive will come from equation (1) which has the known matrices on the left of its terms, but with an entirely parallel procedure the proofs and theorems can be carried out for an $m \times m$ matrix as a solution of the equation where all $A$'s occur on the right of the unknown.

The notation that will be adopted is essentially that employed in the literature.‡ We shall use large italic letters to designate matrices having several columns, but when the matrices have a single column of elements we shall indicate the fact by using script capitals. The elements of the $m \times n$ matrix $A$ are the scalars $a_{ij}$, $i=1, 2, \cdots, m$, $j=1, 2, \cdots, n$, and with the convention just stated we may write

$$A = (\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n),$$

where $\mathcal{A}_j$, $j=1, 2, \cdots, n$, is the matrix formed by the $m$ elements in the $j$th column of $A$. The scalar elements of a matrix will always be indicated by lower case letters.

## Section I

Suppose that the equation (1) has a solution $X$, which of necessity is an $n \times n$ matrix since we are interested particularly in the case where the degree, $p$, of (1) exceeds unity. Let the elementary divisors of $|X - \lambda I|$ be

$$(x_i - \lambda)^{n_i} \qquad (i = 1, 2, \cdots, s),$$

where $\sum_{i=1}^{s} n_i = n$ and where $x_i$, $i=1, 2, \cdots, s$, are not necessarily distinct. Then let

---

* E.Q., pp. 621–631.

† Cecioni, loc. cit.

‡ Bôcher, *Introduction to Higher Algebra*, 1907. Muth, *Theorie und Anwendung der Elementartheiler*.

$$\overline{X} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & X_s \end{pmatrix}$$

and

$$X_i = \begin{pmatrix} x_i & 1 & 0 \cdots 0 & 0 \\ 0 & x_i & 1 \cdots 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 \cdots x_i & 1 \\ 0 & 0 & 0 \cdots 0 & x_i \end{pmatrix} \qquad (i = 1,2, \cdots, s),$$

where $X_i$ is an $n_i \times n_i$ matrix.* The determinant $|\overline{X} - \lambda I|$ will then have the same elementary divisors that $|X - \lambda I|$ has and there consequently exists a non-singular $n \times n$ matrix, $Q$, such that

$$X = Q\overline{X}Q^{-1}. \dagger \tag{2}$$

Hence equation (1) becomes

$$\sum_{j=0}^{p} A_j Q \overline{X}^{p-j} Q^{-1} = 0,$$

or, multiplying through by $Q$ on the right, we have

$$\sum_{j=0}^{p} A_j Q \overline{X}^{p-j} = 0. \tag{3}$$

Let

$$\nu_i = n_1 + n_2 + \cdots + n_{i-1} \qquad (i = 1,2, \cdots, s),$$

where $\nu_1 = 0$, and let $\mathfrak{Q}_j$ be the $m \times 1$ matrix formed by the $j$th column of $Q$. With this we find that the matric equation giving the $n_i$ columns following the first $\nu_i$ of the members of (3) is

$$\sum_{j=0}^{p} A_j (\mathfrak{Q}_{\nu_i+1}, \mathfrak{Q}_{\nu_i+2}, \cdots, \mathfrak{Q}_{\nu_i+n_i}) X_i^{p-j} = 0. \tag{4}$$

Now let

$$A(\lambda) = \sum_{j=0}^{p} A_j \lambda^{p-j} = (a_{hk}(\lambda)),$$

where $\lambda$ is a scalar variable; the element $a_{hk}(\lambda)$, $h = 1, 2, \cdots, m$, $k = 1, 2, \cdots, n$, is a polynomial in $\lambda$ whose degree does not exceed $p$. Also let

* Zero matrices of whatever dimensions will be given by 0; their horizontal and vertical orders will be clear from the expressions in which they occur.

† Bôcher, op. cit., p. 283.

$$A^{(r)}(\lambda) = (a_{hk}^{(r)}(\lambda)),$$

where $a_{hk}^{(r)}(\lambda)$ is the $r$th derivative of $a_{hk}(\lambda)$ with respect to $\lambda$. Since

$$a_{hk}(\lambda) \equiv a_{hk}(x_i) + a'_{hk}(x_i)(\lambda - x_i) + \frac{a_{hk}^{(2)}(x_i)}{2!}(\lambda - x_i)^2 + \cdots + \frac{a_{hk}^{(p)}(x_i)}{p!}(\lambda - x_i)^p,$$

we are then permitted to write

$$A(\lambda) \equiv A(x_i) + A'(x_i)(\lambda - x_i) + \frac{A^{(2)}(x_i)}{2!}(\lambda - x_i)^2 + \cdots + \frac{A^{(p)}(x_i)}{p!}(\lambda - x_i)^p. \tag{5}$$

Now

$$X_i^t = \begin{pmatrix} x_i^t, & \binom{t}{1}x_i^{t-1}, & \cdots, & \binom{t}{n_i-1}x_i^{t-n_i+1} \\ 0, & x_i^t, & \cdots, & \binom{t}{n_i-2}x_i^{t-n_i+2} \\ \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & \cdots, & x_i^t \end{pmatrix}$$

where $t$ is a positive integer and $\binom{t}{j}$ is the $(j+1)$st binomial coefficient. For $X_i^0$ we may write the unit $n_i \times n_i$ matrix. Because the elements $x_i$ are scalars they are commutative with $\mathfrak{Q}_j$, $j=1, 2, \cdots, n$, and we can then readily verify the fact that the first, second, $\cdots$, $n_i$th columns of the members of (4) satisfy the equations

$$\begin{aligned} &A(x_i)\mathfrak{Q}_{\rho_i+1} = 0, \\ &A'(x_i)\mathfrak{Q}_{\rho_i+1} + A(x_i)\mathfrak{Q}_{\rho_i+2} = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\ &\frac{A^{(n_i-1)}(x_i)}{(n_i-1)!}\mathfrak{Q}_{\rho_i+1} + \frac{A^{(n_i-2)}(x_i)}{(n_i-2)!}\mathfrak{Q}_{\rho_i+2} + \cdots + A(x_i)\mathfrak{Q}_{\rho_i+n_i} = 0. \end{aligned} \tag{6}$$

Now by hypothesis $Q$ is a non-singular matrix, hence none of its columns can be a zero matrix. The first equation above is a system of $m$ linear and

homogeneous equations* in the $n$ elements of $\mathfrak{Q}_{\nu_i+1} \neq 0$; we can therefore conclude that the rank, $r_{i1}$, of $A(x_i)$ is less than $n$.† When $m<n$ this condition is always satisfied by an arbitrary $x_i$. We shall henceforth speak of $x_i$, $i=1, 2, \cdots, s$, as the characteristic values of $X$. We therefore have the following theorem.‡

THEOREM I. *The characteristic values, $x_i$, $i=1, 2, \cdots, s$, of $X$, a solution of* (1), *must be such that the rank of $A(x_i)$ is less than $n$.*

When $m<n$ this theorem gives us no information regarding the characteristic values of $X$.

COROLLARY. *The number of distinct characteristic values of $X$ cannot exceed the number of distinct values of $\lambda$ for which the rank of $A(\lambda)$ is less than $n$.*

The rank of $A(x_i)$ being $r_{i1}$, then the number of linearly independent matrices, $\mathfrak{Q}_{\nu_i+1}$, that satisfy the first equation of (6), is $n-r_{i1}$.

THEOREM II. *If $X$ is a solution of* (1) *whose characteristic determinant $|X-\lambda I|$ has the elementary divisors*

$$(x_i - \lambda)^{n_i} \qquad (i = 1,2, \cdots, s),$$

*the number of elementary divisors that are powers of the same linear factor $x_i-\lambda$ cannot exceed $n-r_{i1}$, where $r_{i1}$ is the rank of $A(x_i)$.*

We shall now establish two theorems basing our reasoning on all $n_i$ equations of (6).

## SECTION II

$A(\lambda)$ is an $m \times n$ $\lambda$-matrix for which we shall form an equivalent $\lambda$-matrix, $B(\lambda)$, by means of the following transformations:

1. The interchange of any two columns.

2. The multiplication of each element of a column of $A(\lambda)$ by the same constant not zero.

3. The addition to the elements of one column of the products of corre-

* The rows of the matrix $\sum_{j=0}^{p} \lambda_j A_j$ may be regarded as linearly independent for all values of $\lambda_j$, $j=0, 1, 2, \cdots, p$, for if this matrix has less than $m$ linearly independent rows then the same linearly dependent rows may be dropped from all the matrices $A_j$, $j=0, 1, \cdots, p$, without affecting the solution of (1) on the right in any way.

† Bôcher, op. cit., p. 47.

‡ This theorem for the case $m=n$ was first announced by Sylvester, E.Q., p. 621. He gives an interesting account of the growth of the ideas that led him to its discovery in Nature (Nov. 13, 1884, pp. 35–36). His proof is an intuitive one. A rigorous proof of a theorem slightly more general than Sylvester's was given by H. B. Phillips, *Functions of matrices*, Theorem I, American Journal of Mathematics, vol. 41 (1919), p. 287.

sponding elements of another column by one and the same polynomial in $\lambda$.* The matrix $A(\lambda)Q$ is equivalent to $A(\lambda)$, since $Q$ is a non-singular matrix and the multiplication of $A(\lambda)$ by it on the right employs only the three transformations above. Let

$$R(\lambda) = \begin{pmatrix} R_1(\lambda) & 0 & \cdots & 0 \\ 0 & R_2(\lambda) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & R_s(\lambda) \end{pmatrix},$$

where

$$R_i(\lambda) = \begin{pmatrix} 1 & 0 & 0 \cdots 0 \\ (\lambda - x_i) & 1 & 0 \cdots 0 \\ (\lambda - x_i)^2 & 0 & 1 \cdots 0 \\ \cdots & \cdots & \cdots \\ (\lambda - x_i)^{n_i - 1} & 0 & 0 \cdots 1 \end{pmatrix} \quad (i = 1, 2, \cdots, s).$$

Then $|R(\lambda)| = 1$ and $A(\lambda)QR(\lambda)$ is again equivalent to $A(\lambda)$. If $A(\lambda)$ be replaced by its identically equal given by (5) and if the operations indicated by $A(\lambda)QR(\lambda)$ be performed, we see by (6) that

$$A(\lambda)QR(\lambda) \equiv (B_1(\lambda), B_2(\lambda), \cdots, B_s(\lambda)),$$

where

$$B_i(\lambda) = \begin{pmatrix} (\lambda - x_i)^{n_i} b_{1,\nu_i+1}(\lambda), & b_{1,\nu_i+2}(\lambda), \cdots, & b_{1,\nu_i+n_i}(\lambda) \\ (\lambda - x_i)^{n_i} b_{2,\nu_i+1}(\lambda), & b_{2,\nu_i+2}(\lambda), \cdots, & b_{2,\nu_i+n_i}(\lambda) \\ \cdots & \cdots & \cdots \\ (\lambda - x_i)^{n_i} b_{m,\nu_i+1}(\lambda), & b_{m,\nu_i+2}(\lambda), \cdots, & b_{m,\nu_i+n_i}(\lambda) \end{pmatrix}$$

and the elements $b_{k_i,\nu_i+h}(\lambda)$, $k = 1, 2, \cdots, m$, $h = 1, 2, \cdots, n_i$, are polynomials whose degree in $\lambda$ does not exceed $p$.

THEOREM III. *If $X$ is a solution of* (1) *and if $|X - \lambda I|$ has the elementary divisors $(x_i - \lambda)^{n_i}$, $i = 1, 2, \cdots, s$, then the $\lambda$-matrix, $A(\lambda)$, can be transformed by means of elementary transformations of its columns to an equivalent matrix the elements of whose $(n_1 + n_2 + \cdots + n_{i-1} + 1)$st column, $i = 1, 2, \cdots, s$, have the common factor $(\lambda - x_i)^{n_i}$.*

COROLLARY 1. *If $m = n$, $|A(\lambda)|$ is divisible by the characteristic determinant $\phi(\lambda) = |X - \lambda I|$, of a solution of* (1), *or may be identically zero.*

* Bôcher, op. cit., p. 262.

COROLLARY 2. *If $m>n$ the determinant of an $n\times n$ matrix formed from $A(\lambda)$, by deleting any $m-n$ of its rows, is divisible by the characteristic determinant, $\phi(\lambda)=|X-\lambda I|$, of a solution of* (1), *or is identically zero.*

If $m<n$ no such a simple corollary may be deduced from the theorem; this case will be taken up in Section III.

We shall now seek restrictions upon the degree and upon the number of elementary divisors of $|X-\lambda I|$ corresponding to the base $(x_i-\lambda)$.

Regard only the first $j-1$, $j\leqq n_i$, equations of (6). These may be looked upon as $m(j-1)$ linear and homogeneous equations in the $n(j-1)$ elements of $\mathfrak{Q}_{r_i+k}$, $k=1, 2, \cdots, j-1$, and may be written in the form

$$(7)\qquad \begin{pmatrix} A(x_i) & 0 & \cdots 0 \\ A'(x_i) & A(x_i) & \cdots 0 \\ \dfrac{A^{(2)}(x_i)}{2!} & A'(x_i) & \cdots 0 \\ \cdots & \cdots & \cdots \\ \dfrac{A^{(j-2)}(x_i)}{(j-2)!} & \dfrac{A^{(j-3)}(x_i)}{(j-3)!} & \cdots A(x_i) \end{pmatrix} \begin{pmatrix} \mathfrak{Q}_{r_i+1} \\ \mathfrak{Q}_{r_i+2} \\ \cdot \\ \cdot \\ \cdot \\ \mathfrak{Q}_{r_i+j-1} \end{pmatrix} = 0.$$

Let the rank of the matrix of the coefficients of this system be $r_{i,j-1}$. Since $Q$ is non-singular and $j\leqq n_i$, no $\mathfrak{Q}_{r_i+k}$, $k=1, 2, \cdots, j-1$, is zero, so that $r_{i,j-1}$ must be less than the number of elements in $\mathfrak{Q}_{r_i+k}$, $k=1, 2, \cdots, j-1$; that is,

$$n(j-1)-r_{i,j-1}>0.$$

Suppose we have an entirely general set of matrices $\mathfrak{Q}_{r_i+k}$, $k=1, 2, \cdots, j-1$, satisfying (7), which we shall designate by

$$\mathfrak{Q}_{r_i+k}^{(j-1)} \qquad (k=1,2,\cdots,j-1).$$

Let the number of arbitrary parameters among the elements of $\mathfrak{Q}_{r_i+k}^{(j-1)}$ that are independent of those appearing in the elements of $\mathfrak{Q}_{r_i+h}^{(j-1)}$, $1\leqq h<k$, be given by $e_{ik}^{(j-1)}$. Then the total number of independent parameters that may enter the matrices satisfying (7) is

$$(8)\qquad \sum_{k=1}^{j-1} e_{ik}^{(j-1)} = n(j-1)-r_{i,j-1}.$$

Now regard the system given by the first $j$ equations of (6) in the same way. The most general set of matrices satisfying this system is

$$\mathfrak{Q}_{r_i+k}^{(j)} \qquad (k=1,2,\cdots,j).$$

Let $e_{ik}^{(j)}$ be the number of independent parameters that can occur in $\mathfrak{Q}_{\nu_i+k}^{(j)}$ that are moreover independent of those in any matrix $\mathfrak{Q}_{\nu_i+h}^{(j)}$, $1 \leqq h < k$. Then, as above,

$$\sum_{k=1}^{j} e_{ik}^{(j)} = nj - r_{ij}. \tag{9}$$

We shall show that

$$r_{ij} - r_{i,j-1} < n, \; j \leqq n_i,$$

where we let $r_{i0} = 0$.

Evidently if $\mathfrak{Q}_{\nu_i+k}^{(j-1)}$, $k = 1, 2, \cdots, j-1$, satisfy the first $j-1$ equations of (6) and if $\mathfrak{Q}_{\nu_i+k}^{(j)}$, $k = 1, 2, \cdots, j$, satisfy the corresponding system of $j$ equations, then

$$\mathfrak{Q}_{\nu_i+1}^{(j)}, \quad \mathfrak{Q}_{\nu_i+2}^{(j)} + \mathfrak{Q}_{\nu_i+1}^{(j-1)}, \quad \mathfrak{Q}_{\nu_i+3}^{(j)} + \mathfrak{Q}_{\nu_i+2}^{(j-1)}, \cdots, \quad \mathfrak{Q}_{\nu_i+j}^{(j)} + \mathfrak{Q}_{\nu_i+j-1}^{(j-1)}$$

is again a set satisfying the first $j$ equations of (6) with absolutely no restrictions upon the arbitrary parameters that are in either set; hence we may conclude that

$$e_{ik}^{(j)} \geqq e_{i,k-1}^{(j-1)} \qquad (k = 1, 2, \cdots, j), \tag{10}$$

where we let $e_{i0}^{(j-1)} = 0$. From this fact and from (8) and (9) we can conclude that

$$nj - r_{ij} \geqq n(j-1) - r_{i,j-1}.$$

The equality sign here cannot hold, for if it did then

$$e_{ik}^{(j)} = e_{i,k-1}^{(j-1)} \qquad (k = 1, 2, \cdots, j),$$

and $e_{i1}^{(j)}$ would be zero. That is, $\mathfrak{Q}_{\nu_i+1}^{(j)}$ has no arbitrary parameters, and since $A(x_i)\mathfrak{Q}_{\nu_i+1}^{(j)} = 0$ is a system of linear homogeneous equations in the elements of $\mathfrak{Q}_{\nu_i+1}^{(j)}$ and if no arbitrary parameter occur among them, all must be zero. This is not possible for $j \leqq n_i$, since $Q$ is a non-singular matrix. Hence

$$nj - r_{ij} > n(j-1) - r_{i,j-1},$$

or

$$r_{ij} - r_{i,j-1} < n, \; j \leqq n_i.$$

THEOREM IV. *The degree $n_i$ of the elementary divisor $(x_i - \lambda)^{n_i}$ of the characteristic determinant, $|X - \lambda I|$, of a solution of* (1) *cannot equal nor exceed the least positive integer, $j$, for which*

$$r_{ij} - r_{i,j-1} \geqq n.$$

We can get still more general results. From (8), (9), and (10) we see that

$$e_{i1}^{(j)} \leqq nj - r_{ij} - (n(j-1) - r_{i,j-1}).$$

Now the number of matrices, $\mathfrak{Q}_{r_i+1}^{(k_i)}$, $k_i \geqq j$, that are linearly independent cannot exceed $e_{i1}^{(j)}$, where this is the number of arbitrary parameters in which the elements of $\mathfrak{Q}_{r_i+1}^{(j)}$ are expressed. Corresponding to each elementary divisor $(x_i-\lambda)^{k_i}$, having a given base $(x_i-\lambda)$, we have a matrix $\mathfrak{Q}_{r_i+1}^{(k_i)}$, and since $Q$ is non-singular the number of exponents $k_i$ that exceed $j$ cannot exceed

$$e_{i1}^{(j)} \leqq n - (r_{ij} - r_{i,j-1}).$$

This fact gives us the following important theorem:

THEOREM V. *The number of elementary divisors $(x_i-\lambda)^{n_i}$ of the characteristic determinant of $X$, a solution of* (1), *that are powers of the same linear factor, $x_i-\lambda$, and whose degree equals or exceeds $j$ cannot exceed*

$$n - (r_{ij} - r_{i,j-1}),$$

*where we let $r_{i0}=0$.*

In case $j=1$, we get Theorem II, likewise the preceding theorem may be derived as a special case of the above.

## SECTION III

We shall now take up the study of solutions of (1) when $m<n$ or when the determinants of the $n\times n$ matrices formed from $A(\lambda)$ are all identically zero. The rank, $r$, of $A(\lambda)$, regarded as a $\lambda$-matrix, is then less than $n$. That is, the determinants of all matrices formed by the common elements of more than $r$ rows and columns of $A(\lambda)$ are identically zero, regarding $\lambda$ as a variable, whereas at least one $r\times r$ matrix so formed has as its determinant a polynomial in $\lambda$ not identically zero.* In the present section, we consequently consider the case where $r<n$; hence a solution of (1) may have arbitrary characteristic values, as is evident from considerations leading to Theorem 1.

Suppose that the characteristic determinant of a solution of (1) has the elementary divisors

$$(\rho_i - \lambda)^{m_i} \qquad (i = 1,2, \cdots, \sigma),$$

where $\rho_i$, $i=1, 2, \cdots, \sigma$, are arbitrary save for certain values that will appear later, and the elementary divisors

$$(x_j - \lambda)^{n_j} \qquad (j = 1,2, \cdots, s),$$

* Bôcher, op. cit., p. 262.

where $x_j$, $j=1, 2, \cdots, s$, are constants. In particular, the latter may be certain ones of the exceptional values mentioned above, or those values, if any, of $\lambda$ for which the rank of $A(\lambda)$ is less than $r$. Hence the normal form of this solution is given by

$$\overline{X} = \begin{pmatrix} R_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & R_2 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & R_\sigma & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & X_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & X_s \end{pmatrix},$$

where

$$R_i = \begin{pmatrix} \rho_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \rho_i & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \rho_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \rho_i \end{pmatrix} \qquad (i = 1,2, \cdots, \sigma)$$

is of order $m_i$, and where

$$X_j = \begin{pmatrix} x_j & 1 & 0 & \cdots & 0 & 0 \\ 0 & x_j & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & x_j & 1 \\ 0 & 0 & 0 & \cdots & 0 & x_j \end{pmatrix} \qquad (j = 1,2, \cdots, s)$$

is of order $n_j$ and $\sum_{i=1}^{\sigma} m_i + \sum_{j=1}^{s} n_j = n$.

In this case again a non-singular $n \times n$ matrix, $Q$, must exist such that

$$X = Q\overline{X}Q^{-1},$$

but the elements $q_{hk}$, $h, k=1, 2, \cdots, n$, of $Q$ are now rational functions of the parameters $\rho_i$, $i=1, 2, \cdots, \sigma$; we shall assume, without restricting our problem in any way, that these elements are polynomials in the parameters, $\rho_i$, and that $Q^{-1}$ alone has rational functions as elements not all of which may be polynomials. Let

$$\mu_i = m_1 + m_2 + \cdots + m_{i-1} \qquad (i = 1,2, \cdots, \sigma),$$

and

$$\nu_j = \mu_\sigma + m_\sigma + n_1 + n_2 + \cdots + n_{j-1} \qquad (j = 1,2, \cdots, s),$$

where $\mu_1=0$, $\nu_1=m_\sigma+\mu_\sigma$; then by a procedure entirely parallel to that in Section I, we obtain the equations corresponding to (6) for the $m_i$ columns following the first $\mu_i$ or for the $n_j$ following the first $\nu_j$ of an equation corresponding to (3). Since the sequence in which elements $R_i$ and $X_j$ are written in $\overline{X}$ is immaterial, we shall write the system corresponding to (6) for $i=1$ with the subscript of $\rho$ suppressed. We have then

$$
\begin{aligned}
&A(\rho)\,\mathfrak{Q}_1(\rho) \equiv 0,\\
&A'(\rho)\,\mathfrak{Q}_1(\rho) + A(\rho)\,\mathfrak{Q}_2(\rho) \equiv 0,\\
&\qquad\cdots\cdots\cdots\cdots\cdots,\\
&\frac{A^{(m_1-1)}(\rho)}{(m_1-1)!}\mathfrak{Q}_1(\rho) + \frac{A^{(m_1-2)}(\rho)}{(m_1-2)!}\mathfrak{Q}_2(\rho) + \cdots + A(\rho)\,\mathfrak{Q}_{m_1}(\rho) \equiv 0,
\end{aligned}
\tag{11}
$$

where we are justified in writing $\mathfrak{Q}_k(\rho)$, $k=1, 2, \cdots, m_1$, as matrices of a single column having elements that are polynomials in $\rho$ alone. The other parameters $\rho_2, \rho_3, \cdots, \rho_\sigma$ must drop out identically from these equations because the coefficients, $A^{(k)}(\rho)$, $k=0, 1, \cdots, m_i-1$, are independent of them. Evidently for the constant characteristic values the study reduces to that taken up in the preceding sections.

We can readily show that the first identity of (11),

$$A(\rho)\,\mathfrak{Q}_1(\rho) \equiv 0, \tag{12}$$

is satisfied by only $n-r$ matrices:

$$\mathfrak{Q}_{1j}(\rho) \qquad (j = 1,2, \cdots, n-r),$$

where $r$ is the rank of $A(\lambda)$, and where $\mathfrak{Q}_{1j}(\rho)$ are such that the rank of the matrix

$$(\mathfrak{Q}_{11}(\rho), \mathfrak{Q}_{12}(\rho), \cdots, \mathfrak{Q}_{1,n-r}(\rho))$$

is $n-r$. Every other matrix $\mathfrak{Q}_1(\rho)$ satisfying (12) is expressible linearly in terms of these $n-r$ in the following form:

$$d(\rho)\,\mathfrak{Q}_1(\rho) \equiv \sum_{j=1}^{n-r} d_j(\rho)\,\mathfrak{Q}_{1j}(\rho), \tag{13}$$

where $d(\rho)$ and $d_i(\rho)$, $i=1, 2, \cdots, n-r$, are polynomials in $\rho$.

If $j$ of the parameters $\rho_1, \rho_2, \cdots, \rho_\sigma$ are identical, then $j$ columns of $Q$ must satisfy (12), hence the matrix formed by those $j$ columns will be of rank $j$ only if $j \leqq n-r$.

THEOREM VI. *The number of elementary divisors* $(\rho_i-\lambda)^{m_i}$, $i=1, 2, \cdots, \sigma$, *of the characteristic determinant of a solution of* (1) *having the same base* $(\rho_i-\lambda)$ *or in which the bases* $(\rho_i-\lambda)$, $i=1, 2, \cdots, \sigma$, *may simultaneously be equals, cannot exceed* $n-r$.

Suppose that $\mathfrak{Q}_1(\lambda)=(q_{i1}(\lambda))$ satisfies the identity $A(\lambda)\mathfrak{Q}_1(\lambda)\equiv 0$; then

$$\sum_{j=1}^{n} a_{ij}(\lambda)q_{i1}(\lambda) \equiv 0 \qquad (i = 1,2, \cdots, m), \tag{14}$$

and differentiating with respect to $\lambda$, we have

$$\sum_{j=1}^{n} a'_{ij}(\lambda)q_{j1}(\lambda) + \sum_{j=1}^{n} a_{ij}(\lambda)q'_{j1}(\lambda) \equiv 0 \quad (i = 1,2, \cdots, m),$$

or in matrices*

$$A'(\lambda)\mathfrak{Q}_1(\lambda) + A(\lambda)\mathfrak{Q}_1'(\lambda) \equiv 0,$$

where $A'(\lambda)$ and $\mathfrak{Q}_1'(\lambda)$ are the matrices whose elements are the first derivatives with respect to $\lambda$ of the corresponding elements of $A(\lambda)$ and $\mathfrak{Q}_1(\lambda)$ respectively. Continuing in this way to higher derivatives, and replacing $\lambda$ by $\rho$ in the first $m_1$ equations, including (12), so obtained, we see that

$$\mathfrak{Q}_1(\rho), \mathfrak{Q}_1'(\rho), \frac{\mathfrak{Q}_1^{(2)}(\rho)}{2!}, \cdots, \frac{\mathfrak{Q}_1^{(m_1-1)}(\rho)}{(m_1-1)!}$$

is a set of matrices satisfying (11), where $\mathfrak{Q}_1(\rho)$ is a matrix satisfying the first equation of the system.

Let

$$\mathfrak{Q}_{11}(\rho), \mathfrak{Q}_{12}(\rho), \cdots, \mathfrak{Q}_{1m_1}(\rho)$$

be any $m_1$ matrices satisfying (12), and let

$$\mathfrak{Q}_1(\rho), \mathfrak{Q}_2(\rho), \cdots, \mathfrak{Q}_{m_1}(\rho)$$

be matrices satisfying (11), and suppose

$$\mathfrak{Q}_1(\rho) \equiv \mathfrak{Q}_{11}(\rho),$$

$$\mathfrak{Q}_j(\rho) \equiv \mathfrak{S}_j(\rho) + \frac{\mathfrak{Q}_{11}^{(j-1)}(\rho)}{(j-1)!} \qquad (j = 2,3, \cdots, m_1),$$

where $\mathfrak{S}_j(\rho)$ will have a column of $n$ polynomial elements. By substitution in (11) we see that

$$\mathfrak{S}_j(\rho) \qquad (j = 2,3, \cdots, m_1)$$

* Frobenius, L.S., p. 16. Schlesinger, *Vorlesungen über lineare Differentialgleichungen*, 1908, p. 22.

must satisfy a system like (11) with the last equation omitted, and that $\mathfrak{S}_2(\rho)$ must satisfy the first equation. Hence we may let

$$\mathfrak{S}_2(\rho) = \mathfrak{Q}_{12}(\rho),$$

and suppose

$$\mathfrak{S}_j(\rho) \equiv \mathfrak{T}_j(\rho) + \frac{\mathfrak{Q}_{12}^{(j-2)}(\rho)}{(j-2)!} \qquad (j = 3,4, \cdots, m_1);$$

then $\mathfrak{T}_3(\rho)$ must again satisfy (12). Continuing in this way we find that

$$\mathfrak{Q}_{11}(\rho),\ \mathfrak{Q}'_{11}(\rho) + \mathfrak{Q}_{12}(\rho),\ \frac{\mathfrak{Q}_{11}^{(2)}(\rho)}{2!} + \mathfrak{Q}'_{12}(\rho) + \mathfrak{Q}_{13}(\rho),$$

$$\cdots,\ \frac{\mathfrak{Q}_{11}^{(m_1-1)}(\rho)}{(m_1-1)!} + \frac{\mathfrak{Q}_{12}^{(m_1-2)}(\rho)}{(m_1-2)!} + \cdots + \mathfrak{Q}_{1m_1}(\rho)$$

is the general set of matrices satisfying (11), where $\mathfrak{Q}_{1j}(\rho)$, $j=1, 2, \cdots, m_1$, are general matrices satisfying only the first. The matrix of which these are the first, second, third, $\cdots$ and $m_1$st column we shall designate by $Q_{m_1}(\rho_1)$; similarly for the remaining elementary divisors of $|X-\lambda I|$ we have

$$Q_{m_i}(\rho_i) \qquad (i = 2,3, \cdots, \sigma),$$

where $Q_{m_i}(\rho_i)$ is a matrix of $m$ rows and $m_i$ columns.

THEOREM VII. *If*

$$X = Q\overline{X}Q^{-1}$$

*is a solution of* (1) *and* $\overline{X}$ *is its normal form, if* $(\rho_i-\lambda)^{m_i}$, $i=1, 2, \cdots, \sigma$, *are the elementary divisors of the characteristic determinant of* $X$, *and if* $\mathfrak{Q}_{1j}(\lambda)$, $j=1, 2, \cdots, m_1$, *are matrices satisfying the identity*

$$A(\lambda)\mathfrak{Q}(\lambda) \equiv 0, \tag{12}$$

*then the* $m_i$ *columns following the first* $m_1+m_2+\cdots+m_{i-1}$ *of* $Q$ *are given by*

$$Q_{m_i}(\rho_i) \equiv \left(\mathfrak{Q}_{11}(\rho_i), \mathfrak{Q}'_{11}(\rho_i) + \mathfrak{Q}_{12}(\rho_i), \cdots, \frac{\mathfrak{Q}_{11}^{(m_i-1)}(\rho)}{(m_i-1)!} + \frac{\mathfrak{Q}_{12}^{(m_i-2)}(\rho_i)}{(m_i-2)!} + \cdots + \mathfrak{Q}_{1m_i}(\rho_i)\right)$$

*in which the matrices* $\mathfrak{Q}_{1j}$, $j=1, 2, \cdots, m_i$, *are taken arbitrarily from those satisfying* (12), *and the values of* $\rho_i$, $i=1, 2, \cdots, \sigma$, *that must be excepted are the values of these parameters for which the rank of*

$$Q = (Q_{m_1}(\rho_1), Q_{m_2}(\rho_2), \cdots, Q_{m_\sigma}(\rho_\sigma))$$

*is less than* $n$.

A sufficient condition that (1) have a solution is that $Q$, $|Q|$ not identically zero, may be constructed from the solutions of (12) in the manner given above. To do this may be impossible for solutions whose characteristic determinants have all elementary divisors of the kind $(\rho_i-\lambda)^{m_i}$, $i=1, 2, \cdots, \sigma$, but (1) may have a solution in whose characteristic determinant the elementary divisors $(x_j-\lambda)^{n_j}$, $j=1, 2, \cdots, s$, where $x_j$ are constants, are permitted. This possibly arises when the determinants of all $r\times r$ matrices formed from $A(\lambda)$ have the greatest common factor $p(\lambda)$, not a constant. Let the zeros of $p(\lambda)$ be $x_j$, $j=1, 2, \cdots, s$; for these values the rank, $r_{j1}$, of $A(x_j)$ is less than $r$, the rank of $A(\lambda)$. The columns that $Q$ must have permitting the characteristic values $x_j$, $j=1, 2, \cdots, s$, must satisfy the equations (6) of the preceding sections. If the columns so determined for $x_j$, $j=1, 2, \cdots, s$, and the columns satisfying (11) permit us to build up a non-singular matrix, $Q$, then a solution of (1) whose characteristic determinant has the corresponding elementary divisors

$$(\rho_i-\lambda)^{m_i} \quad (i=1,2,\cdots,\sigma), \qquad (x_j-\lambda)^{n_j} \quad (j=1,2,\cdots,s),$$

exists. This is clarified by an example that will be given later. We can get a rather simple sufficient condition for the existence of solutions of (1) all of whose characteristic values are variables.

Suppose that $\mathfrak{Q}(\lambda)$ such that

$$A(\lambda)\mathfrak{Q}(\lambda)\equiv 0$$

and whose elements $q_i(\lambda)$, $i=1, 2, \cdots, n$, are polynomials of degree $q$ in $\lambda$, exists. Then

$$\mathfrak{Q}(\lambda)\equiv\mathfrak{Q}(\rho)+\mathfrak{Q}'(\rho)(\lambda-\rho)+\frac{\mathfrak{Q}^{(2)}(\rho)}{2!}(\lambda-\rho)^2+\cdots+\frac{\mathfrak{Q}^{(q)}(\rho)}{q!}(\lambda-\rho)^q,$$

where $\mathfrak{Q}^{(k)}(\lambda)$, $k=0, 1, \cdots, q$, is the matrix whose elements are the $k$th derivatives of corresponding elements of $\mathfrak{Q}(\lambda)$, is an identity in $\lambda$ and $\rho$. This identity may be written matricly as

$$\mathfrak{Q}(\lambda)\equiv\left(\mathfrak{Q}(\rho),\mathfrak{Q}'(\rho),\cdots,\frac{\mathfrak{Q}^{(q)}(\rho)}{q!}\right)\begin{pmatrix}1\\ \lambda-\rho\\ (\lambda-\rho)^2\\ \vdots\\ (\lambda-\rho)^q\end{pmatrix}.$$

If here we replace $\lambda$ by $\rho_i$, $i=1, 2, \cdots, n$, respectively, and regard the matrix, $Q$, having $\mathfrak{Q}(\rho_i)$, $i=1, 2, \cdots, n$, as its columns, we have

$$(15)\qquad \begin{aligned} Q &\equiv (\mathfrak{Q}(\rho_1), \mathfrak{Q}(\rho_2), \cdots, \mathfrak{Q}(\rho_n)) \\ &\equiv \left(\mathfrak{Q}(\rho), \mathfrak{Q}'(\rho), \cdots, \frac{\mathfrak{Q}^{(q)}(\rho)}{q!}\right) \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \rho_1-\rho & \rho_2-\rho & \cdots & \rho_n-\rho \\ (\rho_1-\rho)^2 & (\rho_2-\rho)^2 & \cdots & (\rho_n-\rho)^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ (\rho_1-\rho)^q & (\rho_2-\rho)^q & \cdots & (\rho_n-\rho)^q \end{pmatrix}, \end{aligned}$$

in which the first factor of the right member is an $n\times(q+1)$ matrix and the second is a $(q+1)\times n$ matrix. The rank of $Q$ cannot exceed the rank of either factor; and if $q<n-1$, the rank of $Q$, whatever the parameters $\rho_i$, $i=1, 2, \cdots, n$, may be is less than $n$. On the other hand, if $q$ equals or exceeds $n-1$, the first factor will certainly be of rank $n$ if the elements of $\mathfrak{Q}(\rho)$ are linearly independent, because the determinant of the matrix formed by its first $n$ columns is the Wronskian of those elements multiplied by a constant not zero. The second factor will be of rank $n$ if and only if $\rho_i\neq\rho_j$, $i\neq j$, and $q\geqq n-1$, for the determinant of the matrix formed by its first $n$ rows is the determinant of Vandermonde.

Now if the elements of $\mathfrak{Q}(\lambda)$ are linearly independent and if $\rho_i\neq\rho_j$, $i\neq j$, then $Q$ is of rank $n$, for regarding the first factor of the right member of (15) as a $\lambda$-matrix with $\rho$ as the variable parameter and the second factor as elementary transformations, as defined earlier in the present section, upon the columns of the first, we see that this must be the case.*

With an appropriate change in the argument above, we can show that the matrix

$$Q = (Q_1, Q_2, \cdots, Q_\sigma),$$

where

$$Q_i = \left(\mathfrak{Q}(\rho_i), \mathfrak{Q}'(\rho_i), \cdots, \frac{\mathfrak{Q}^{(m_i-1)}(\rho_i)}{(m_i-1)!}\right),$$

and $\rho_i\neq\rho_j$ for $i\neq j$, is of rank $n$ if the elements of $\mathfrak{Q}(\lambda)$ are linearly independent.†

---

* The degree, $q$, of the elements of $\mathfrak{Q}(\lambda)$ must equal or exceed $n-1$, if they are linearly independent.

† See Eriksson, *Sur une extension du déterminant de Vandermonde*, Arkiv för Matematik, Astronomi, och Fysik, vol. 16 (1922), No. 13.

THEOREM VIII. *A sufficient condition that equation* (1) *have a solution* $X$ *whose characteristic determinant has the elementary divisors*

$$(\rho_i - \lambda)^{m_i} \qquad (i = 1, 2, \cdots, \sigma),$$

*where* $\rho_i$, $i=1, 2, \cdots, \sigma$, *are all distinct but otherwise arbitrary parameters and where* $\sum_{i=1}^{\sigma} m_i = n$, *is that the equation*

$$A(\lambda)\mathfrak{Q}(\lambda) \equiv 0$$

*be satisfied by a matrix* $\mathfrak{Q}(\lambda)$, *whose elements* $q_j(\lambda)$, $j=1, 2, \cdots, n$, *are linearly independent. This condition is both necessary and sufficient for the existence of a solution of* (1) *whose characteristic determinant has the single elementary divisor* $(\rho-\lambda)^n$, *where* $\rho$ *is an arbitrary parameter.*

This theorem may be extended in case the identity $A(\lambda)\mathfrak{Q}(\lambda)\equiv 0$ is satisfied by $k$ matrices, $\mathfrak{Q}_k(\lambda)$, $k \leqq n-r$, such that

$$\sum_{j=1}^{k} b_j(\lambda)\mathfrak{Q}_j(\lambda) \equiv 0$$

cannot be satisfied by polynomials $b_i(\lambda)$, not all identically zero, and such that the elements $q_{ij}(\lambda)$, $i=1, 2, \cdots, n$, of $\mathfrak{Q}_j(\lambda)$ are linearly independent, to permit $k$ of the elementary divisors $(\rho_i-\lambda)^{m_i}$ to have the same base $(\rho_i-\lambda)$ or to take equal values.

Since the equation

$$A(\lambda)\mathfrak{Q}(\lambda) \equiv 0 \tag{16}$$

plays such an important part in case the rank of $A(\lambda)$ is less than $n$, we shall take up briefly a method of determining $\mathfrak{Q}(\lambda)$ whose elements $q_i(\lambda)$, $i=1, 2, \cdots, n$, are polynomials in $\lambda$ of degree not exceeding $j$. Such a matrix will always exist if $j$ is taken sufficiently large, and is given by

$$\mathfrak{Q}(\lambda) \equiv \mathfrak{Q}_0 + \mathfrak{Q}_1\lambda + \mathfrak{Q}_2\lambda^2 + \cdots + \mathfrak{Q}_j\lambda^j,$$

where $\mathfrak{Q}_k$, $k=0, 1, \cdots, j$, are matrices having a single column of $n$ constants, $q_{ik}$, $i=1, 2, \cdots, n$, $k=0, 1, 2, \cdots, j$, and where

$$q_i(\lambda) = q_{i0} + q_{i1}\lambda + \cdots + q_{ij}\lambda^j \qquad (i = 1, 2, \cdots, m);$$

not all $q_{i0}$, $i=1, 2, \cdots, n$, are zeros. Now

$$A(\lambda) = A_0\lambda^p + A_1\lambda^{p-1} + \cdots + A_p,$$

and because (16) must be satisfied for all values of $\lambda$, we have at once the equations*

* Compare Kronecker, *Vorlesungen über die Theorie der Determinanten*, vol. 1, 1903, p. 363.

$$
\begin{aligned}
&A_p\mathfrak{Q}_0 = 0,\\
&A_{p-1}\mathfrak{Q}_0 + A_p\mathfrak{Q}_1 = 0,\\
&\cdots\cdots\cdots,\\
&A_0\mathfrak{Q}_0 + A_1\mathfrak{Q}_1 + \cdots + A_j\mathfrak{Q}_j = 0,\\
&A_0\mathfrak{Q}_1 + \cdots + A_{j-1}\mathfrak{Q}_j = 0,\\
&\cdots\cdots\cdots,\\
&A_0\mathfrak{Q}_j = 0.
\end{aligned}
\tag{17}
$$

This is a system of $(p+j+1)m$ linear and homogeneous equations in the $(j+1)n$ coefficients of the elements of $\mathfrak{Q}(\lambda)$. The degree $j$ must be taken sufficiently large so that the rank of the matrix of the coefficients is less than $(j+1)n$. This is always possible since the rank, $r$, of $A(\lambda)$ is less than $n$. In fact, $j$ can be so chosen that the rank of the system does not exceed $nj+r$, then the general solution of (17) gives us a fundamental set of polynomials $\mathfrak{Q}_j(\lambda)$, $j=1, 2, \cdots, n-r$, in terms of which all others satisfying (12) may be expressed linearly with polynomial coefficients. Moreover, from (17) it is evident that if either $A_0$ or $A_p$ is of rank $n$ then $\mathfrak{Q}(\lambda)$ satisfying (16) must be identically zero, hence *equation* (1) *may have solutions having variable characteristic values only when the rank of* $A_0$ *and of* $A_p$ *is less than* $n$.

We shall illustrate the above theory with two examples:

EXAMPLE 1.

$$(1,0,1)X^2 + (-3,2,-2)X + (0,2,-3) = 0.$$

Here

$$A(\lambda) = (\lambda^2 - 3\lambda, 2\lambda + 2, \lambda^2 - 2\lambda - 3),$$

and according to equations (17) a solution of the first degree in $\lambda$ of (12) is

$$\mathfrak{Q}(\lambda) = q_0\begin{pmatrix}\lambda+1\\0\\-\lambda\end{pmatrix} + q_1\begin{pmatrix}0\\\lambda-3\\-2\end{pmatrix},$$

where $q_0$ and $q_1$ are arbitrary constants. The general solution of (12) is then

$$b_0(\lambda)\begin{pmatrix}\lambda+1\\0\\-\lambda\end{pmatrix} + b_1(\lambda)\begin{pmatrix}0\\\lambda-3\\-2\end{pmatrix},$$

where $b_0(\lambda)$ and $b_1(\lambda)$ are polynomials in $\lambda$. If we let $b_0(\lambda)\equiv\lambda$ and $b_1(\lambda)\equiv 1$ we have

$$\mathfrak{Q}(\lambda) = \begin{pmatrix}\lambda^2+\lambda\\\lambda-3\\-\lambda^2-2\end{pmatrix},$$

a matrix satisfying (12) whose elements are linearly independent. Hence by Theorem VIII the given equation has the following solutions:

$$X_1 = Q_1\begin{pmatrix} \rho & 1 & 0 \\ 0 & \rho & 1 \\ 0 & 0 & \rho \end{pmatrix}Q_1^{-1}, \quad Q_1 = \begin{pmatrix} \rho^2+\rho, & 2\rho+1, & 1 \\ \rho-3, & 1, & 0 \\ -\rho^2-2, & -2\rho, & -1 \end{pmatrix};$$

$$X_2 = Q_2\begin{pmatrix} \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \\ 0 & 0 & \rho_2 \end{pmatrix}Q_2^{-1}, \quad Q_2 = \begin{pmatrix} \rho_1^2+\rho_1, & 2\rho_1+1, & \rho_2^2+\rho_2 \\ \rho_1-3, & 1, & \rho_2-3 \\ -\rho_1^2-2, & -2\rho_1, & -\rho_2^2-2 \end{pmatrix},$$

$$\rho_1 \neq \rho_2;$$

$$X_3 = Q_3\begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{pmatrix}Q_3^{-1}, \quad Q_3 = \begin{pmatrix} \rho_1^2+\rho_1, & \rho_2^2+\rho_2, & \rho_3^2+\rho_3 \\ \rho_1-3, & \rho_2-3, & \rho_3-3 \\ -\rho_1^2-2, & -\rho_2^2-2, & -\rho_3^2-2 \end{pmatrix},$$

$$(\rho_1-\rho_2)(\rho_2-\rho_3)(\rho_3-\rho_1) \neq 0;$$

however, according to Theorem VI elementary divisors having the same base may occur, at most, twice in the characteristic determinants of solutions, since the rank of $A(\lambda)$ is unity. Such solutions are

$$X_4 = Q_4\begin{pmatrix} \rho_1 & 1 & 0 \\ 0 & \rho_1 & 0 \\ 0 & 0 & \rho_2 \end{pmatrix}Q_4^{-1},$$

where

$$Q_4 = \begin{pmatrix} \rho_1+1 & 1 & 0 \\ 0 & 0 & \rho_2-3 \\ -\rho_1 & -1 & -2 \end{pmatrix}, \quad \text{or} \quad Q_4 = \begin{pmatrix} 0 & 0 & \rho_2+1 \\ \rho_1-3 & 1 & 0 \\ -2 & 0 & -\rho_2 \end{pmatrix};$$

$\rho_2$ in the first cannot equal 3, in the second $-1$;

$$X_5 = Q_5\begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & \rho_3 \end{pmatrix}Q_5^{-1},$$

$$Q_5 = \begin{pmatrix} \rho_1+1 & 0 & \rho_3+1 \\ 0 & \rho_2-3 & 0 \\ -\rho_1 & -2 & -\rho_3 \end{pmatrix}, \quad \begin{matrix} \rho_1 \neq \rho_3, \\ \rho_2 \neq 3, \end{matrix}$$

or

$$Q_5 = \begin{pmatrix} 0 & \rho_2+1 & 0 \\ \rho_1-3 & 0 & \rho_3-3 \\ -2 & -\rho_2 & -2 \end{pmatrix}, \quad \begin{matrix} \rho_1 \neq \rho_3, \\ \rho_2 \neq -1; \end{matrix}$$

the last two solutions, $X_4$ and $X_5$, are less restrictive than are $X_2$ and $X_3$.

EXAMPLE 2.

$$\begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix} X^2 + \begin{pmatrix} -3 & 1 & -3 \\ 3 & -1 & -1 \end{pmatrix} X + \begin{pmatrix} 2 & 2 & 2 \\ -2 & -2 & 1 \end{pmatrix} = 0.$$

This equation has no solutions all of whose characteristic values are variables, since the equation $A(\lambda)\mathfrak{Q}(\lambda) \equiv 0$ is satisfied only by

$$\mathfrak{Q}(\lambda) = b(\lambda) \begin{pmatrix} \lambda + 2 \\ -\lambda^2 + 3\lambda - 2 \\ 0 \end{pmatrix},$$

whose elements are not linearly independent. Theorem VIII is not satisfied; however the rank of $A(\lambda)$ is unity when $\lambda = 1$, and for this value of $\lambda$ we have the following equations corresponding to (5):

$$\begin{pmatrix} 0 & 3 & 0 \\ 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = 0,$$

$$\begin{pmatrix} -1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} + \begin{pmatrix} 0 & 3 & 0 \\ 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = 0,$$

$$\begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} + \begin{pmatrix} -1 & 1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} + \begin{pmatrix} 0 & 3 & 0 \\ 0 & -3 & 0 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = 0.$$

The first of these is satisfied by

$$\mathfrak{Q}_1 = \begin{pmatrix} q_1 \\ 0 \\ q_3 \end{pmatrix};$$

the first two by

$$\mathfrak{Q}_1 = \begin{pmatrix} 3r_2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathfrak{Q}_2 = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix};$$

all three by

$$\mathfrak{Q}_1 = \begin{pmatrix} 3r_2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathfrak{Q}_2 = \begin{pmatrix} 4r_2 + 3s_2 \\ r_2 \\ 0 \end{pmatrix}, \quad \mathfrak{Q}_3 = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix},$$

where the constants are arbitrary. The given equation then has solutions:

$$X_1 = Q_1\begin{pmatrix} \rho & 1 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & 1 \end{pmatrix}Q_1^{-1},\quad Q_1 = \begin{pmatrix} \rho+2 & 1 & q_1 \\ -\rho^2+3\rho-2 & -2\rho+3 & 0 \\ 0 & 0 & q_3 \end{pmatrix},$$

$$q_3(\rho^2+4\rho-8) \neq 0\ ;$$

$$X_2 = Q_2\begin{pmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}Q_2^{-1},\quad Q_2 = \begin{pmatrix} \rho_1+2 & \rho_2+2 & q_1 \\ -\rho_1^2+3\rho_1-2 & -\rho_2^2+3\rho_2-2 & 0 \\ 0 & 0 & q_3 \end{pmatrix},$$

$$q_3(\rho_1-\rho_2)(\rho_1\rho_2+2\rho_1+2\rho_2-8) \neq 0\ ;$$

$$X_3 = Q_3\begin{pmatrix} \rho & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}Q_3^{-1},\quad Q_3 = \begin{pmatrix} \rho+2 & 3r_2 & r_1 \\ -\rho^2+3\rho-2 & 0 & r_2 \\ 0 & 0 & r_3 \end{pmatrix},$$

$$r_2r_3(\rho^2-3\rho+2) \neq 0\ ;$$

$$X_4 = Q_4\begin{pmatrix} \rho & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}Q_4^{-1},\quad Q_4 = \begin{pmatrix} \rho+2 & q_1 & r_1 \\ -\rho^2+3\rho-2 & 0 & 0 \\ 0 & q_3 & r_3 \end{pmatrix},$$

$$(q_1r_3-q_3r_1)(\rho^2-3\rho+2) \neq 0\ ;$$

$$X_5 = Q_5\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}Q_5^{-1},\quad Q_5 = \begin{pmatrix} 3r_2 & 4r_2+3s_2 & s_1 \\ 0 & r_2 & s_2 \\ 0 & 0 & s_3 \end{pmatrix},$$

$$r_2s_3 \neq 0\ ;$$

where $\rho_1, \rho_2, q_1, q_3, r_1, r_2, r_3, s_1, s_2, s_3$ are entirely arbitrary save that the determinant of $Q$ in each case shall be different from zero. The given equation has no solutions whose characteristic determinants have as elementary divisors

$$(-2 \pm 2\cdot 3^{1/2} - \lambda)^2,\ (1-\lambda),$$
$$(2-\lambda),\ (1-\lambda)^2,$$
$$2-\lambda,\ 1-\lambda,\ 1-\lambda,$$

because for each of these groups equations (6) lead to a singular $Q$, nor as elementary divisors

$$1-\lambda,\ 1-\lambda,\ 1-\lambda$$

because of Theorem II.

UNIVERSITY EXTENSION DIVISION,
UNIVERSITY OF WISCONSIN,
MILWAUKEE, WIS.

# ON THE ZEROS OF LINEAR PARTIAL FRACTIONS*

BY

MORRIS MARDEN†

## I. INTRODUCTION

**1. Problem and method.‡** The present paper owes its origin to a suggestion from Professor Walsh§ that I try to carry to a correct completion Professor Coble's analytical proof‖ of the following theorem due to the former:¶

---

* Presented to the Society, December 27, 1928; received by the editors in January, 1929.

† National Research Fellow.

‡ Bibliography. The following articles are cited at least twice. Each will hereafter be referred to by the letter B and a roman numeral.

BI: M. Bôcher, Proceedings of the American Academy of Arts and Sciences, vol. 40 (1904), pp. 469–484.

BII: J. L. Walsh, Comptes Rendus du Congrès International des Mathématiciens, Strasbourg, 1920, pp. 1–4.

BIII: J. L. Walsh, these Transactions, vol. 19 (1918), pp. 291–298.

BIV: J. L. Walsh, these Transactions, vol. 22 (1921), pp. 101–116.

BV: J. L. Walsh, these Transactions, vol. 24 (1922), pp. 31–69.

BVI: J. L. Walsh, Proceedings of the National Academy of Sciences, vol. 8 (1922), pp. 139–141.

BVII: J. L. Walsh, Rendiconti del Circolo Matematico di Palermo, vol. 46 (1922), pp. 1–13.

BVIII: A. B. Coble, Bulletin of the American Mathematical Society, vol. 27 (1921), pp. 434–437.

BIX: M. Marden, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 363–370.

§ I wish to express my deep gratitude to Professor Walsh for his many suggestions and criticisms, and, above all, his constant encouragement in the course of this work.

‖ This theorem was proved geometrically by Walsh (see BIV). Subsequently an analytical treatment of the same problem was published by Coble (see BVIII). Although suggestive as to the manipulation of the difficult algebra of the problem, the latter article is open to the following objections. (1) It quotes Walsh's theorem incorrectly by substituting "the interior of a circle" for "the circular region." (2) It considers in the proof only the case of the regions $C_j$ being the interiors of circles. (3) It implicitly assumes that the locus is finite whenever the regions $C_j$ are finite. This need not be the case, as is shown in chapter IV §2 of the present paper. (4) It implicitly assumes without proof that the locus is a simply-connected region. Walsh's theorem, as stated by Coble, was later challenged by T. Nakahara (Tôhoku Mathematical Journal, vol. 23 (1924), p. 97) as if it were Walsh's own statement. Nakahara's contribution to the subject consists of a set of simple conditions for the locus of the theorem to be finite. The conditions given are, however, only sufficient, not both necessary and sufficient, as announced by Nakahara. They may be deduced more easily by the methods of chapter IV §2 of the present paper.

¶ See BIV, pp. 101–112. In Walsh's statement of the theorem, the point $z$ is defined by means of the real constant cross ratio $(z\, z_1\, z_2\, z_3) = \lambda$. Our definition is, however, equivalent to his, provided $\lambda \neq 0, \infty$.

By a circular region is meant in the theorem the interior and circumference of a circle, the exterior and circumference of a circle, a half-plane including its boundary line, or an entire plane.

*If the points $z_1$, $z_2$, $z_3$ vary independently and have circular regions as their respective loci, then the locus of the point $z$ defined by the equation*

$$\frac{m_1}{z-z_1}+\frac{m_2}{z-z_2}+\frac{m_3}{z-z_3}=0, \tag{1}$$

*where $m_1$, $m_2$, $m_3$ are real constants whose sum is zero, is also a circular region.*

In the course of my attempts in this direction, I found that the methods needed in the study of this problem were sufficiently powerful for an investigation of the zeros of the more general partial fraction

$$F=\sum_{j=1}^{q+1}\frac{m_j}{z-z_j},$$

where the $m_j$ are real constants whose sum is or is not zero. This investigation leads to a wide generalization, after the manner of Walsh, of theorems due to Gauss, Lucas, Bôcher and Walsh,* concerning the location of the zeros of the derivative of a rational function and of the zeros of the jacobian of two binary forms. For instance, its results enable one to locate approximately the positions of the zeros of the derivative of the polynomial

$$f(z)=f_1(z)f_2(z)\cdots f_k(z)$$

when all the zeros of each factor polynomial $f_j(z)$ are known to lie in or on a circle $C_j$. The approximation given is furthermore the best possible, in the sense that the geometrical limits involved are actually attained.

In place of merely proving the above theorem, I therefore set out to solve the following problem: *To find the locus $R$ of the zeros of the partial fraction $F$ when the points $z_j$ vary independently and have circular regions $C_j$ as their respective loci.*†

The answer to this general problem turns out to be a reasonable generalization of the above theorem. In place of a circular region as the locus of the zeros of $F$ comes what we shall call a *$p$-circular domain.*‡ By a $p$-circular domain we shall mean either the entire plane or a set of simply-connected

* See chapters III and V for statements of these theorems.

† Two interpretations of the problem are the following:

(a). Suppose a particle of mass $m_j$ is placed at each point $z_j$,—a particle which repels every other particle according to the inverse distance law. In the field of force so created, the zeros of $F$ are the points of equilibrium. Our problem is to find the whereabouts of these points of equilibrium, when it is known that each point $z_j$ may occupy any position within its circular region $C_j$. (Cf. BI, p. 475.)

(b). Given the function $z=f(z_1, z_2, \cdots, z_{q+1})$ defined by the equation $F=0$ in a certain chunk of $(q+1)$-dimensional complex space. Our problem is to find the image of this chunk upon the $z$-plane.

‡ Cf. BV, footnote p. 39, where Walsh guesses the result for the locus $R$, and BV, p. 38–43, where he derives some necessary properties of the locus $R$.

closed regions which contain all the ovals of a given $p$-circular $2p$-ic curve and which are bounded only by ovals of this $2p$-ic curve.* For example, if $p$ were equal to two and the bicircular quartic consisted of one oval enclosed in a second, then as a bicircular domain we might have the entire plane, the interior and boundary of the outer oval, or the exterior and boundary of the inner oval, but not the ring-shaped region between the two ovals.

More specifically, we shall find the locus $R$ to be a $p$-circular domain, where

$$p = q - 1 + \lim_{s \to 0} \left( \frac{n}{n+s}, \right) \text{ and } n = \sum_{j=1}^{q+1} m_j.$$

This domain, if not the entire plane, is bounded by part or all of the $p$-circular $2p$-ic curve with the equation†

$$\sum_{j=1}^{q+1} \frac{nm_j}{C_j} - \sum_{j=1, k=j+1}^{q+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} = 0,$$

where $C_j = 0$ is the equation of the circle bounding the circular region $C_j$, and where $\Delta_{jk}$ is, according to a definite rule, the square of either the external or internal common tangent of the circles $C_j$ and $C_k$.

As our way to this result is rather long, it is perhaps worthwhile here to outline our method of attack and to point out its principal difficulties.

We shall obtain the above information about the locus $R$ by successively determining certain two-dimensional envelopes.‡ If $R_k$ denotes the locus of the zeros of $F$ when $z_1, z_2, z_3, \cdots, z_k$ vary over their respective circular regions but the points $z_{k+1}, z_{k+2}, \cdots, z_{q+1}$ are fixed, then $R_k$ may also be regarded as the envelope of the loci $R_{k-1}$ when $z_k$ describes as locus the circular region $C_k$. Thus, by computing successively the envelopes $R_0, R_1, \cdots, R_{q+1}$, we finally arrive at the locus $R$.

This method of successive envelopes entails as principal steps (1) the algebraic computation of the equations of the possible boundary curves of each locus $R_k$; (2) from these possible boundary curves the choice of the actual boundary curve of each locus $R_k$; and (3) the proof that each locus $R_k$ is composed only of simply-connected regions.

---

* We mean by a $p$-circular $2p$-ic curve one which is represented by an equation of the form

$$a(x^2 + y^2)^p + f(x,y) = 0$$

where $f(x, y)$ is a polynomial of degree less than $2p$. If $a \neq 0$, the curve intersects the line at infinity (in the projective plane) $p$ times in each circular point at infinity and therefore cannot meet this line in any additional points. In other words, if $a \neq 0$, the curve consists entirely of closed ovals.

† The entire left-hand side of the equation given above lacks, of course, the factor $\prod_{j=1}^{q+1} C_j$. This factor will be omitted quite often in this paper, for the purpose of simplifying the formulas.

‡ This method is suggested in the footnote of BIV, p. 102.

As a gradual approach to the difficulties involved, we shall consider at first two special cases: in chapter II, a case (case I) where all the $m_j$ are positive real numbers and the regions $C_j$ are finite, and in chapter IV, case II, where the $m_j$ are real numbers whose sum is zero, but where the regions $C_j$ are arbitrary. The advantage of this approach is that the work once done for case I in connection with steps (1) and (3) holds almost equally for case II and that therefore in case II we are left free to overcome the difficulties of step (2) caused by the arbitrariness of the numbers $m_j$ and the regions $C_j$. Furthermore, from the results of case II, we shall readily be able to infer the results in the most general case.

Having thus found the locus $R$, we shall try to verify our results as including those of Walsh and, finally in a closing chapter, we shall present two generalizations, the one for the zeros of the expression $F+\lambda$ where $\lambda$ is a constant, real or complex, and the other for the zeros of $F$ when the $m_j$ are complex constants.

2. **Notation.** We now proceed to introduce a system of abridged notation which will enable us expeditiously to perform the first of the above-described steps in the determination of the loci $R_k$. This system is essentially the same as that used by Coble in BVIII. Like him, we employ the symbol $[C_1C_2]$ for the invariant

$$2(A_1A_2 + B_1B_2) - b_1a_2 - b_2a_1 \equiv a_1a_2(R_1^2 + R_2^2 - d_{12}^2), \tag{1.21}$$

which frequently appears in the computations. In the identity (1.21) $R_j$ is the radius of the circle $C_j$, $d_{12}$ is the distance between the centers of the circles $C_1$ and $C_2$ and the other letters are the coefficients in the equation

$$C_j \equiv a_j(x^2 + y^2) - 2A_jx - 2B_jy + b_j = 0,$$

$a_j$, $A_j$, $B_j$, $b_j$ all being real numbers. If $C_j$ is written in the form

$$C_j \equiv a_jz\bar{z} - \bar{\beta}_jz - \beta_j\bar{z} + b_j = 0,$$

where $z=x+iy$ and $\beta_j=A_j+iB_j$, then

$$[C_1C_2] = \beta_1\bar{\beta}_2 + \bar{\beta}_1\beta_2 - a_1b_2 - a_2b_1.$$

As may be easily verified, the quantity $[C_1\ C_2]$ obeys the following laws of combination:

1 $$[C_2C_1] = [C_1C_2],$$

2 $$[C_1\ kC_2] = k[C_1C_2],$$

3 $$[C_1\ C_2 + k] = [C_1C_2] - ka_1,$$

4 $$[C_1\ C_2 + C_3] = [C_1C_2] + [C_1C_3],$$

5 $$[C_1 + kC_2\ C_1 + kC_2] = [C_1C_1] + 2k[C_1C_2] + k^2[C_2C_2],$$

where $k$ is an ordinary constant.

When $C_j$ degenerates into the null circle

$$P_j = (x - x_j)^2 + (y - y_j)^2,$$

additional laws are

$$[C_j P_k] = -C_{jk}, \tag{6}$$

$$[P_k P_j] = -P_{jk} = -P_{kj}, \tag{7}$$

where $C_{jk} = C_j(z_k)$ and $P_{jk} = P_j(z_k)$. These latter quantities have the properties that, if the point $P_k$ is on the circle $C_j$, then $C_{jk} = 0$; if $P_k \equiv P_j$, then $P_{jk} = 0$; and always

$$P_{jk} + P_{jl} - P_{kl} \equiv (z_j - z_k)(\bar{z}_j - \bar{z}_l) + (\bar{z}_j - \bar{z}_k)(z_j - z_l).$$

In terms of the above notation, the radius of the circle $C_j$ is $R_j^2 = [C_j C_j]/(2a_j^2)$ and the cosine of the angle between the two circles $C_j$ and $C_k$ is $\cos\ (C_j,\ C_k) = [C_j C_k]/\{[C_j C_j]\ [C_k C_k]\}^{1/2}$. As necessary and sufficient conditions we therefore have

(1) for $C_j$ to be a null circle: $[C_j\ C_j] = 0$;

(2) for $C_j$ and $C_k$ to be orthogonal: $[C_j\ C_k] = 0$; and

(3) for $C_j$ and $C_k$ to be mutually tangent:

$$[C_j C_k]^2 - [C_j C_j][C_k C_k] = 0. \tag{1.22}$$

The factors of the left hand side of (1.22), the invariants

$$\sigma_{jk} \equiv \frac{1}{a_j a_k}\{[C_j C_j]^{1/2}[C_k C_k]^{1/2} - [C_j C_k]\}$$

$$\equiv d_{jk}^2 - (R_j - R_k)^2,$$

and

$$\tau_{jk} \equiv -\frac{1}{a_j a_k}\{[C_j C_j]^{1/2}[C_k C_k]^{1/2} + [C_j C_k]\} \equiv d_{jk}^2 - (R_j + R_k)^2,$$

are plainly, if positive, the squares of the external and internal common tangents respectively of the circles $C_j$ and $C_k$, and therefore must reduce to

$$\sigma_{jk} = \tau_{jk} = C_{kj}$$

when $C_j \equiv P_j$ and to

$$\sigma_{jk} = \tau_{jk} = P_{jk}$$

when both $C_j \equiv P_j$ and $C_k \equiv P_k$.

## II. Special case I: Positive numbers $m_j$. Finite circular regions $C_j$

**1. The locus $R_0$.** The locus $R_0$ (cf. definition in I § 1) consists merely of the $q$ zero-points of the partial fraction $F$. Since $n \neq 0$ and the points $z_j$ are finite, the points comprising $R_0$ are also finite.

The locus $R_0$ may be regarded as the null $q$-circular $2q$-ic curve

$$E_0 \equiv \left\{ \prod_1^{q+1} (z - z_j)(\bar{z} - \bar{z}_j) \right\} \left( \sum_1^{q+1} \frac{m_j}{z - z_j} \right) \left( \sum_1^{q+1} \frac{m_j}{\bar{z} - \bar{z}_j} \right) = 0.$$

To write the latter expression entirely in terms of the notation of chapter I § 2, we observe that, when $q = 1$,

$$\left( \frac{m_1}{z - z_1} + \frac{m_2}{z - z_2} \right) \left( \frac{m_1}{\bar{z} - \bar{z}_1} + \frac{m_2}{\bar{z} - \bar{z}_2} \right) = \frac{m_1^2}{P_1} + \frac{m_2^2}{P_2}$$

$$+ \frac{m_1 m_2}{P_1 P_2} \{ (z - z_1)(\bar{z} - \bar{z}_2) + (\bar{z} - \bar{z}_1)(z - z_2) \} = \frac{n m_1}{P_1} + \frac{n m_2}{P_2} - \frac{m_1 m_2 P_{12}}{P_1 P_2}.$$

This suggests what can be proved by mathematical induction; namely, that

$$E_0 \equiv \left( \prod_1^{q+1} P_j \right) \left\{ \sum_1^{q+1} \frac{n m_j}{P_j} - \sum_{\substack{j=1 \\ k=j+1}}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k} \right\}.$$

**2. The locus $R_1$.** The locus $R_1$, evidently a bounded point set, may be considered as the map of the region $C_1$ (the locus of the point $z_1$) by means of the transformation $z = \psi(z_1)$ defined by the equation $F = 0$. As the derivative

$$\frac{dz}{dz_1} = \frac{m_1}{(z - z_1)^2} \left[ \sum_1^{q+1} \frac{m_j}{(z - z_j)^2} \right]^{-1} \neq 0, \infty ,$$

the transformation is conformal and consequently $R_1$ is a set of one or more bounded two-dimensional continua.

If $B_1$ denotes the boundary curve of $R_1$ and $z$ is any point of $B_1$, there must correspond to $z$ as a point of $R_1$ at least (actually only) one point $z_1$ in the region $C_1$. This point $z_1$ must lie *on* the circle $C_1$. For, were $z_1$ inside the circle $C_1$, a small neighborhood of $z_1$ could be mapped by the function $\psi(z_1)$ upon a neighborhood of $z$ consisting only of points of $R_1$. The point $z$ could under those circumstances not be a point of $B_1$, contrary to the hypothesis. Consequently, when $z_1$ traces the circle $C_1$, at least part of the resulting locus of the point $z$ is $B_1$.

The last fact enables us to obtain the equation of $B_1$ if we observe that, though representing in the variable $z$ a $q$-circular $2q$-ic, the equation $E_0 = 0$ represents in the variable $z_1$ a null circle $G_0$. The null circle $G_0$ is, in fact,

the point $z_1$ corresponding to a given point $z$. If $z$ is a point of $B_1$, by what has been proved above, $G_0$ lies on the circle $C_1$ and therefore*

$$[E_0 C_1]_{z_1} = 0 . \tag{2.21}$$

Computed according to the rules given in I § 2, the left hand side of (2.21) is

$$E_1 \equiv C_1 \prod_2^{q+1} P_j \left\{ \sum_2^{q+1} \frac{n m_j}{P_j} - \sum_{\substack{j=2 \\ k=j+1}}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k} \right\} + \prod_2^{q+1} P_j \left\{ n m_1 - \sum_2^{q+1} \frac{m_1 m_j C_{1j}}{P_j} \right\}.$$

Every point $z$ of $B_1$ must therefore be a point of the curve $E_1 = 0$.

As is clear from the leading term of its equation, the curve $E_1 = 0$ is always a $q$-circular $2q$-ic and as such consists exclusively of closed ovals. Among these ovals we may discern two types: the first, ovals $O_1$ which neither contain nor are contained in any other oval of $E_1 = 0$; the second, ovals $O_2$ not belonging to the first type.

The facts that $O_1$ forms part of $R_1$, that $K_1$ is a set of finite two-dimensional regions, and that $B_1$ is part of the curve $E_1 = 0$ certify to the entire interior of $O_1$'s belonging to $R_1$. Whether or not the part of $R_1$ associated with an $O_2$ is simply-connected, needs however further investigation. For the present, we shall assume it also to be a simply-connected region (cf. II § 6).

The locus $R_1$ is therefore the interior and boundary of all the ovals of the $q$-circular $2q$-ic curve $E_1 = 0$.

3. **The locus $R_2$.** Regarded as the envelope of the loci $R_1$, the locus $R_2$ consists also of a set of finite two-dimensional regions.

If $z$ is a point of $B_2$, the boundary curve of $R_2$, to $z$ corresponds (via the equation $F = 0$) at least one pair of points $(z_1, z_2)$ in the regions $C_1$, $C_2$ respectively. This pair must actually lie on the circles $C_1$, $C_2$. For, were, for example, $z_1$ inside the circle $C_1$, the point $z_2$ could be considered fixed and the arguments of II § 2 used to contradict the hypothesis that the point $z$ is on $B_2$.

In other words, if the points $z_1$ and $z_2$ trace their respective circles simultaneously, the resulting locus of the point $z$ will include the curve $B_2$. That is to say, $B_2$ is part of the envelope of the curves $B_1$ or $E_1 = 0$ when the point $z_2$ describes as locus the circle $C_2$.

From this fact comes the equation of $B_2$ if we notice that, in the variable $z_2$, the equation $E_1 = 0$ represents a circle $G_1$. The circle $G_1$ corresponding to a point $z$ on $B_2$ must consist of at least one point on the circle $C_2$ and of no points

* The subscript $z_1$ means that in this operation $z_1$ is considered as the independent variable.

inside the circle $C_2$; in other words, it must be tangent to $C_2$. A necessary condition on $z$ as a point of $B_2$ is therefore

$$[E_1C_2]^2_{z_2} - [C_2C_2]_{z_2}[E_1E_1]_{z_2} = 0. \tag{2.31}$$

Computed by the rules of I § 2, the terms of the left hand side of this expression become

$$-[E_1C_2] = \left(C_1C_2\prod_3^{q+1}P_j\right)\left\{\sum_1^2\frac{nm_j}{C_j}+\sum_3^{q+1}\frac{nm_j}{P_j}+\frac{m_1m_2[C_1C_2]}{C_1C_2}\right.$$
$$\left.-\sum_{j=1,k=3}^{j=2,k=q+1}\frac{m_jm_kC_{jk}}{C_jP_k}-\sum_{j=3,k=j+1}^{q+1}\frac{m_jm_kP_{jk}}{P_jP_k}\right\};$$
$$[E_1E_1] = m_1^2m_2^2\left(\prod_3^{q+1}P_j\right)^2[C_1C_1].$$

Equation (2.31) accordingly factors into the two equations

$$E_2 \equiv -[E_1C_2] - m_1m_2\prod_3^{q+1}P_j[C_1C_1]^{1/2}[C_2C_2]^{1/2} = 0,$$

$$E_2' \equiv -[E_1C_2] + m_1m_2\prod_3^{q+1}P_j[C_1C_1]^{1/2}[C_2C_2]^{1/2} = 0,$$

at least one of which every point of $B_2$ must satisfy.

Both equations, having as leading term $n^2(x^2+y^2)^q$, represent $q$-circular $2q$-ic curves. If $(\xi, \eta)$ is a point of the second, i.e. $E_2'(\xi, \eta)=0$, then

$$E_2(\xi,\eta) = -2m_1m_2[C_1C_1]^{1/2}[C_2C_2]^{1/2}\prod_3^{q+1}P_j(\xi,\eta) \leqq 0. \tag{2.32}$$

The significance of the inequality (2.32) becomes clear if we observe that the continuous function $E_2(x, y)$ grows positively infinite as the point $(x, y)$ recedes indefinitely in any direction. The inequality (2.32) therefore expresses a sufficient condition for $(\xi, \eta)$ to lie inside or on an oval of $E_2=0$, i.e., for every oval of $E_2'=0$ to be nested in an oval of $E_2=0$.

If now we assume (cf. II § 6) that $R_2$ is also composed only of simply-connected parts, we obtain $R_2$ as the interior and boundary of all the ovals of the $q$-circular $2q$-ic

$$E_2 = \left(C_1C_2\prod_3^{q+1}P_j\right)\left\{\sum_1^2\frac{nm_j}{C_j}+\sum_3^{q+1}\frac{nm_j}{P_j}-\frac{m_1m_2\sigma_{12}}{C_1C_2}\right.$$
$$\left.-\sum_{j=1,k=3}^{j=2,k=q+1}\frac{m_jm_kC_{jk}}{C_jP_k}-\sum_{j=3,k=j+1}^{q+1}\frac{m_jm_kP_{jk}}{P_jP_k}\right\} = 0.$$

4. **The locus $R_N$.** We may now venture to suppose that $R_N$ consists of the interior and boundary of all the ovals of

$$E_N = \Big(\prod_1^N C_j \prod_{N+1}^{q+1} P_j\Big)\Big\{\sum_1^N \frac{nm_j}{C_j} + \sum_{N+1}^{q+1} \frac{nm_j}{P_j} - \sum_{j=1,k=j+1}^{N} \frac{m_j m_k \sigma_{jk}}{C_j C_k} - \sum_{j=1,k=N+1}^{j=N,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} - \sum_{j=N+1,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k}\Big\} = 0.$$

On the principle of mathematical induction, let us test the correctness of this assumption by determining $R_{N+1}$.

Considerations similar to those of II § 3 lead to the equation

$$[E_N C_{N+1}]^2_{zN+1} - [E_N E_N]_{zN+1}[C_{N+1}C_{N+1}]_{zN+1} = 0 \tag{2.41}$$

as a necessary condition that a point $z$ lie on $B_{N+1}$, the boundary curve of $R_{N+1}$. Deciphered under the rules of I § 2, the terms in (2.41) may be written as

$$-[E_N C_{N+1}] = \prod_1^{N+1} C_j \prod_{N+2}^{q+1} P_j \Big\{ \sum_1^{N+1} \frac{nm_j}{C_j} + \sum_{N+2}^{q+1} \frac{nm_j}{P_j} + \sum_1^N \frac{m_{N+1}m_j[C_{N+1}C_j]}{C_{N+1}C_j} - \sum_{j=1,k=N+2}^{j=N+1,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} - \sum_{j=N+2,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k} - \sum_{j=1,k=j+1}^{N} \frac{m_j m_k \sigma_{jk}}{C_j C_k}\Big\},$$

$$[E_N E_N] = \Big\{\prod_1^N C_j \prod_{N+2}^{q+1} P_j m_{N+1} \sum_1^N \frac{m_j [C_j C_j]^{1/2}}{C_j}\Big\}^2.$$

Every point of $B_{N+1}$ must therefore lie on at least one of the two curves

$$\begin{aligned} E_{N+1} &\equiv -[E_N C_{N+1}] - [E_N E_N]^{1/2}[C_{N+1}C_{N+1}]^{1/2} = 0 \\ &\equiv \prod_1^{N+1} C_j \prod_{N+2}^{q+1} P_j \Big\{ \sum_1^{N+1} \frac{nm_j}{C_j} + \sum_{N+2}^{q+1} \frac{nm_j}{P_j} - \sum_{j=1,k=j+2}^{N+1} \frac{m_j m_k \sigma_{jk}}{C_j C_k} \\ &\quad - \sum_{j=1,k=N+2}^{j=N+1,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} - \sum_{j=N+2,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k}\Big\} = 0; \end{aligned}$$

$$\begin{aligned} E^{(N+1)}_{N+1} &\equiv -[E_N C_{N+1}] + [E_N E_N]^{1/2}[C_{N+1}C_{N+1}]^{1/2} = 0 \\ &\equiv \prod_1^{N+1} C_j \prod_{N+2}^{q+1} P_j \Big\{ \sum_1^{N+1} \frac{nm_j}{C_j} + \sum_{N+2}^{q+1} \frac{nm_j}{P_j} - \sum_{j=1,k=j+1}^{N+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} \\ &\quad - \sum_{j=1,k=N+2}^{j=N+1,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} - \sum_{j=N+2,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k}\Big\} = 0, \end{aligned} \tag{2.42}$$

where $\Delta_{jk} = \tau_{jk}$ if $i$ or $j = N+1$, but $\Delta_{jk} = \sigma_{jk}$ otherwise.

Let us now bring into play the fact that $B_{N+1}$ must be independent of the

order in which the $z_j, j=1, 2, \cdots, N+1$, have begun to vary. If, for instance, the orders of $z_\kappa$ and $z_{N+1}$ are interchanged, the condition upon $z$ as a point of $B_{N+1}$ must be revised so as to read that $z$ satisfies at least one of the two equations $E^{(\kappa)'}_{N+1}=0$ and $E^{(\kappa)}_{N+1}=0$. These new equations may be derived from the old (2.42) by permuting the $z_\kappa$ and $z_{N+1}$. As the form $E_{N+1}$ is symmetric in the subscripts $j=1, 2, \cdots, N+1$, $E^{(\kappa)'}_{N+1}\equiv E_{N+1}$, but

$$E^{(k)}_{N+1} \equiv \prod_1^{N+1} C_j \prod_{N+2}^{q+1} P_j \Big\{ \sum_1^{N+1} \frac{nm_j}{C_j} + \sum_{N+2}^{q+1} \frac{nm_j}{P_j} - \sum_{j=1,k=j+1}^{N+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} - \sum_{j=1,k=N+2}^{j=N+1,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} - \sum_{j=N+2,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k} \Big\} = 0,$$

where $\Delta_{ij}=\tau_{ij}$ if $i$ or $j=\kappa$, but $\Delta_{ij}=\sigma_{ij}$ otherwise.

Let us assume that not all of curves $E^{(k)}_{N+1}=0$ have an arc in common. If $\zeta$ is a point on the curve $E^{(1)}_{N+1}=0$ and $\nu$ a sufficiently small arc of $E^{(1)}_{N+1}=0$ containing $\zeta$, there will exist at least one curve $E^{(\kappa)}_{N+1}=0$ which does not contain $\nu$. Let us further suppose that $\zeta$ is not on the curve $E_{N+1}=0$. If $\zeta$ lies on $B_{N+1}$, it must satisfy at least one of the two equations in each of the sets

(1) $E_{N+1}=0,\ E^{(1)}_{N+1}=0$; and (2) $E_{N+1}=0,\ E^{(\kappa)}_{N+1}=0.$

Yet $\zeta$ fails to satisfy either equation of set (2). It follows therefore from (1) that every point of $B_{N+1}$ is a point of the curve $E_{N+1}=0$.

Up to this point we have supposed not all the curves $E^{(k)}_{N+1}=0$ to have an arc in common. That such is the general state of affairs is evident from the distinctness in the forms $E^{(k)}_{N+1}$. In case, however, the $E^{(k)}_{N+1}=0$ all have an arc in common, we shall use the following reasoning. We shall consider the manifold $M$ consisting of all possible instances of special case I and its submanifold $M'$ composed of all instances for which all the curves $E^{(k)}_{N+1}=0$ have an arc in common. The manifold $M$ is of dimensions $3(N+1)$, the variables being the centers and radii of the $N+1$ circles $C_j$. If the dimensions of $M'$ were less than $3(N+1)$, it would be possible to draw in $M$ a continuous arc which terminates in any chosen point of $M'$ but does not pass through any other points of $M'$. It would accordingly be possible to pass to the case where the curves $E^{(k)}_{N+1}=0$ have a common arc, continuously through cases where this is not so. As the equation $E_{N+1}$ has coefficients depending continuously upon those of the equations $C_j=0$, the curve $B_{N+1}$ would consist, even in the case of the curves $E^{(k)}_{N+1}=0$ having a common arc, only of ovals of the curve $E_{N+1}=0$.

We have then but to prove that the dimensions of $M'$ are less than $3(N+1)$. As the curves $E^{(k)}_{N+1}=0$ all have a common arc, any point of this

arc has coördinates satisfying equations

$$\Gamma^{(s,t)} \equiv E_{N+1}^{(s)} - E_{N+1}^{(t)}$$
$$\equiv 2\left(\prod_1^{N+1} C_j \prod_{N+2}^{q+1} P_j\right)\left\{\frac{m_t[C_tC_t]^{1/2}}{C_t} - \frac{m_s[C_sC_s]^{1/2}}{C_s}\right\}$$
$$\cdot\left\{\sum_1^{N+1}\frac{m_j[C_jC_j]^{1/2}}{C_j} - \frac{m_s[C_sC_s]^{1/2}}{C_s} - \frac{m_t[C_tC_t]^{1/2}}{C_t}\right\} = 0,$$

where $s \neq t$ and $s, t = 1, 2, \cdots, N+1$. The common arc is thus an arc of a circle $Q$ with various distinct equations of the form $\sum \lambda_j C_j = 0$. Equating the several expressions for the center and radius of $Q$, we obtain the equations

$$\frac{\sum_1^{N+1}\lambda_j\alpha_j}{\sum_1^{N+1}\lambda_j} = \frac{\sum_1^{N+1}\mu_j\alpha_j}{\sum_1^{N+1}\mu_j},$$

$$\frac{\left[\sum_1^{N+1}\lambda_jC_j, \sum_1^{N+1}\lambda_jC_j\right]}{\left(\sum_1^{N+1}\lambda_j\right)^2} = \frac{\left[\sum_1^{N+1}\mu_jC_j, \sum_1^{N+1}\mu_jC_j\right]}{\left(\sum_1^{N+1}\mu_j\right)^2}.$$

As we assumed $M'$ to exist, these equations are compatible and define $M'$ as a manifold of dimensions lower than $3(N+1)$.

In every case, therefore, $B_{N+1}$ is made up only of ovals of the curve $E_{N+1}=0$. If we assume the locus $R_{N+1}$ to consist of simply-connected parts, it follows that $R_{N+1}$ is the interior and boundary of all the ovals of the curve $E_{N+1}=0$, a result which certifies to the correctness of the assumption concerning the locus $R_N$.

5. **The locus $R \equiv R_{q+1}$. Its properties.** By setting $N=q+1$ we finally conclude that *the locus R is the interior and the boundary of all the ovals of the q-circular 2q-ic curve**

* The following theorem also results by a limiting process from the work of the preceding sections: *If $\alpha_j'(j=1, 2, \cdots, p)$ are the zeros of the partial fraction*

$$\frac{m_j}{z-\alpha_j}$$

*and $\zeta_m$ and $\zeta_M$ respectively the smallest and largest zeros of the partial fraction*

$$\sum_{j+1}^{q+1}\frac{m_j}{z-\beta_j}$$

*where $\beta_j = \Re(\alpha_j)$, then*

$$\zeta_m \leqq \Re(\alpha_j') \leqq \zeta_M \qquad (j=1, 2, \cdots, p).$$

This is Linfield's theorem for the case $k=1$. See reference BV, p. 54, and B. Z. Linfield, these Transactions, vol. 25 (1923), p. 257.

$$\sum_{1}^{q+1} \frac{nm_j}{C_j} - \sum_{j=1,k=j+1}^{q+1} \frac{m_j m_k \sigma_{jk}}{C_j C_k} = 0. \tag{2.51}$$

As equation (2.51) represents a circular curve, it is of interest to locate the singular foci of this curve. Proceeding in the usual way, we find the circular line through the point $(u, v)$,

$$y = ix + v - iu, \tag{2.52}$$

to intersect curve (2.51) in points with abscissas which satisfy the equation

$$2^q x^q \left\{ \prod_{1}^{q+1}(w - \alpha_j) \sum_{1}^{q+1} \frac{1}{w - \alpha_j} \right\} + \text{terms of lower degree} = 0,$$

where $u+iv=w$ and $a_j+ib_j=\alpha_j$. Line (2.52) will then be asymptotic to curve (2.51), and the point $(u, v)$ a singular focus of curve (2.51) if and only if $(u, v)$ is such that

$$\sum_{1}^{q+1} \frac{1}{w - \alpha_j} = 0.$$

*The singular foci are hence the zeros of F for the $z_j=\alpha_j$*; or, what amounts to the same, *the points of equilibrium in a field of force created by a particle of mass $m_j$ at each $\alpha_j$ repelling according to the inverse distance law.* (See I § 1, footnotes.)*

For further geometric characterization of the singular foci when the $m_j$ are integers, we may appeal to the theorem recently proved by Linfield that "the roots of the derivative of the rational function $f(z)=\Pi_1^n (z-z_j)^{\mu_j}$ are the multiple roots of $f(z)$ to one lower order and the foci of a curve of class $n-1$ which touches each segment $z_j z_k$ $(j \neq k = 1, 2, \cdots, n)$ at the points dividing it in the ratio $\mu_j$ to $\mu_k$." Our singular foci are therefore the ordinary foci of Linfield's curve.†

From the form of the equation (2.51) we can deduce two more facts about the locus $R$.

First, as is obvious, *any point through which pass at least three of the circles $C_j$ is a point of curve* (2.51) *and therefore a point of the locus R.*

Secondly, *the common points of any two of the regions $C_j$ belong to the locus R.*

For suppose a point $P$ were to lie in just two of the regions $C_j$, for definiteness, $C_1$ and $C_2$. Entirely within the regions $C_1$ and $C_2$ let us draw respectively

* The singular foci are therefore points of the locus $R$. The bicircular quartic
$$(x^2 + y^2 + 37)^2 = 64(x^2 + 25)$$
which has its singular foci at the points $(\pm 4, 0)$, points outside of its solitary oval, is an example of a $p$-circular $2p$-ic which, on account of the theorem, cannot serve as the boundary curve of a locus $R$.

† B. Z. Linfield, Bulletin of the American Mathematical Society, vol. 27 (1920), p. 17, and these Transactions, vol. 25 (1923), pp. 239–258.

new circles $C_1'$ and $C_2'$ which pass through the point $P$. The point $P$ is now such that

$$C_1'(P) = C_2'(P) = 0$$

and $C_j(P) > 0$ for all $j \neq 1$ or 2. The expression

$$E' \equiv C_1'C_2' \prod_3^{q+1} C_j \left\{ \sum_1^2 \frac{nm_j}{C_j'} + \sum_3^{q+1} \frac{nm_j}{C_j} - \frac{m_1m_2\sigma_{12}'}{C_1'C_2'} - \sum_{j=1,k=3}^{j=2,k=q+1} \frac{m_jm_k\sigma_{jk}'}{C_j'C_k} - \sum_{j=3,k=j+1}^{q+1} \frac{m_jm_k\sigma_{jk}}{C_jC_k} \right\}$$

formed at $P$ has then the value

$$E'(P) = -m_1m_2\sigma_{12}'C_3(P)C_4(P) \cdots C_{q+1}(P).$$

As neither one of the circles $C_1'$ and $C_2'$ is contained in the other, $\sigma_{12}' > 0$ and therefore $E'(P) < 0$. This means (cf. II § 3) that $P$ lies within the ovals of the $q$-circular $2q$-ic curve $E' = 0$ and accordingly is a point of the locus $R'$, the locus $R$ in which the circles $C_1'$ and $C_2'$ respectively replace the circles $C_1$ and $C_2$. But, as $R'$ is included in the $R$ corresponding to all the original circles $C_j$, the point $P$ is contained in $R$.

If $P$ were assumed to be a common point of just three of the regions $C_j$, for example, $C_1$, $C_2$, $C_3$, we could draw inside each of these three regions new circles $C_1'$, $C_2'$, $C_3'$ which pass through the point $P$. The point $P$ would then be a point of $R''$, the locus $R$ in which the new circles $C'$ replace the old, and therefore a point of the locus $R$ corresponding to all of the original circles $C_j$. Similar arguments hold for common points of four or more of the regions $C_j$.

6. **Simple-connectivity.** There remains for the completion of our study of special case I only to establish the fact that each locus $R_k$ is composed of simply-connected regions. This we shall do for $k = 2$, a typical case. We draw in $R_2$ an arbitrary simple closed curve $K$ and show that, without ever intersecting $B_2$, the curve $K$ can be shrunk continuously to a point in $R_2$.

Considered as the envelope of the $R_1$, the locus $R_2$ consists of two overlapping parts: a part $S_2$ composed of the interior points of the $R_1$ corresponding to $z_2 = \alpha_2$, and a part $T_2$ composed of the closed areas swept out by the moving curve $B_1$, i.e., composed exclusively of points through which passes at least one curve $B_1$.

The curve $K$ is thus composed (1) entirely of points in $S_2$, (2) entirely of points in $T_2$, or (3) partly of points in $T_2$ and partly of points not in $T_2$. If we assume the $R_1$ to be of simply-connected parts, there is nothing further to prove for curves $K$ of type (1). On the same assumption, it is possible

so to deform a curve of type (3) continuously that it becomes a curve of type (2). The proof of the simple-connectivity therefore rests upon the behavior of curves $K$ of the second type.

Let $K$ then be a curve of the second type and $w$ an arbitrary point on it. Through $w$ as a point of $T_2$ passes at least one curve $B_1$, i.e., the equation

$$E_1(w, z_2) = 0 \tag{2.61}$$

has at least one solution $z_2$ in the region $C_2$ and therefore the circle $G_1$ given by (2.61) has at least one point in the region $C_2$. We shall select one of these points and label it $w_2$.

Let us mark the point $w = W$ on $K$, to which will correspond the circle $G_1 = \Lambda_1$ and the point $w_2 = W_2$ on $\Lambda_1$. Beginning at $w = W$, as $w$ traces $K$ continuously clockwise, the circle $G_1$ alters continuously its center and radius beginning with those of $\Lambda_1$. With $G_1$ we so shift $w_2$ as to cause it to trace a continuous curve $\Gamma_2$ which begins at $w_2 = W_2$. When $w$ finally returns to $W$, $G_1$ becomes again $\Lambda_1$ but $w_2$ ends at a point $w_2 = W_2'$ on $\Lambda_1$, not necessarily the same as $W_2$. To make $\Gamma_2$ a closed curve, we add to it an arc $\lambda_2 = W_2W_2'$ of circle $\Lambda_1$, an arc lying in the region $C_2$.

In the same way it is possible to associate with $K$ a curve $\Gamma_1$ in the region $C_1$. We consider again any point $w$ on $K$ and the curve $B_1$ through it. As $B_1$ is the envelope of the curves $E_0 = 0$ when $z_1$ traces the circle $C_1$, there passes through $w$ at least one curve $E_0 = 0$; i.e., the equation

$$E_0(w, w_2, z_1) = 0 \tag{2.62}$$

has at least one solution $z_1$ on the circle $C_1$ and therefore the null circle $G_0$ given by (2.62) lies on the circle $C_1$. We shall denote this $G_0$ by $w_1$; in particular, we shall label the points corresponding to $w = W$ and $w_2 = W_2$, and $w = W$ and $w_2 = W_2'$ by $w_1 = W_1$ and $w_1 = W_1'$ respectively.

As the point $w$ and the corresponding point $w_2$ trace $K$ and $\Gamma_2$ from $W$ and $W_2$ to $W$ and $W_2'$, the point $w_1$ describes on the circle $C_1$ a curve $\Gamma_1$ from $W_1$ to $W_1'$. If the point $w$ is now held fast at $W$ and the point $w_2$ traces $\lambda_2$ from $W_2'$ to $W_2$, the point $w_1$ traces an arc $\lambda_1$ of the circle $C_1$, which arc closes $\Gamma_1$.

So far we have been concerned with selecting, from the infinitude of pairs $(z_1, z_2)$ leading to a point $w$ on $K$, just those which together form closed curves $\Gamma_1$ and $\Gamma_2$ in the regions $C_1$ and $C_2$. The correspondence thus set up between points $w$ of $K$ and $w_1$ and $w_2$ of $\Gamma_1$ and $\Gamma_2$ is one-to-one except that to the point $W$ corresponds all the points on the arcs $\lambda_1$ and $\lambda_2$. The correspondence between points $w_1$ of $\Gamma_1$ and $w_2$ of $\Gamma_2$ is one-to-one without exception.

From now on we shall consider $K$ as generated by the point $z=w$ when $w_2$ and its corresponding $w_1$ trace out $\Gamma_2$ and $\Gamma_1$ simultaneously. By suitably deforming $\Gamma_2$ and $\Gamma_1$, we shall try to shrink $K$ to a point in $R_2$.

Continuous deformations of $\Gamma_1$ and $\Gamma_2$ subject to the following conditions will preserve the essential features of the above-formed correspondences between $w$, $w_1$ and $w_2$: (1) the points $W_1$ and $W_2$ are to remain stationary; (2) points $w_2$ on $\lambda_2$ are to go into points $w_2'$ also on $\lambda_2$, and the corresponding point $w_1$ into the null circle $E_0(W, w_2', w_1)=0$; and, finally, (3) the curves $\Gamma_1$ and $\Gamma_2$ are always to be closed curves lying in the regions $C_1$ and $C_2$.

Subject to these restrictions, let us modify the shapes of $\Gamma_1$ and $\Gamma_2$ slightly, causing $w_1$ and $w_2$ to go into such points $w_1'$ and $w_2'$ of $\Gamma_1'$ and $\Gamma_2'$ that

$$| w_2 - w_2' | < \delta, \quad | w_1 - w_1' | < \delta.$$

A continuous function of $w_1$ and $w_2$, $w$ is carried into $w'$ where

$$| w - w' | < \epsilon, \tag{2.63}$$

and $\epsilon$ is a positive number arbitrarily small for sufficiently small $\delta$.

Of what points the curve $K'$, the transform of $K$, consists may be seen if the point $w_2'$ and the corresponding point $w_1'$ are allowed to trace continuously complete circuits on $\Gamma_2'$ and $\Gamma_1'$, beginning at the points $w_2'=W_2$ and $w_1'=W_1$. The curve $K'$ will be a continuous curve with $W$ and $W'$ as terminal points, $W'$ being another root of the equation $F=0$ where $z_1=W_1$ and $z_2=W_2$.

If $W\neq W'$, the equation $F=0$ then has $r\geqq 2$ distinct roots: $Z_1=W$, $Z_2=W'$, $Z_3, \cdots, Z_r$. For these a number $g(W_1, W_2)$ always exists so that

$$| Z_i - Z_j | \geqq g > 0, \quad i \neq j ;$$

in particular,

$$| W - W' | \geqq g. \tag{2.64}$$

On the other hand, $W'$ may be thought of as the point $w'$ into which $W$ has gone as result of the above-described deformations. From inequality (2.63) it then follows that

$$| W - W' | < \epsilon. \tag{2.65}$$

For $\delta$ sufficiently small, we may take $\epsilon=g/2$ causing inequalities (2.64) and (2.65) to contradict one another. Consequently, $W=W'$: the new $K'$ is a closed curve.

As total effect of our deformations on $\Gamma_1$ and $\Gamma_2$, the curve $K$ has therefore changed into another continuous closed curve $K'$ in $R_2$ which passes through the point $W$. If now we contract $\Gamma_1$ and $\Gamma_2$ continuously to the

points $W_1$ and $W_2$, the curve $K$ will, without ever cutting $B_2$, shrink continuously to the point $W$, a point in $R_2$.

In a similar way we may prove that all the other loci $R_k$ are composed of simply-connected parts. The curve $\Gamma_k$ is chosen as a closed curve entirely in the region $C_k$, but the other auxiliary curves $\Gamma_{k-p}$ as closed curves traced on the circles $C_{k-p}$ by the points of tangency of certain circles $G_{k-p-1}$.

## III. Applications

**1. The derivative of a polynomial.** The results of the previous chapter will now be applied to the location of the zeros of the derivative of the polynomial:

$$f(z) = \prod_1^n (z - z_j) = \prod_1^{q+1} f_j(z),$$

where $f_j$ is a polynomial of degree $m_j$. As

$$f'(z) = f(z)\left\{\sum_1^n \frac{1}{z - z_j}\right\}, \tag{3.1}$$

the zeros of $f'(z)$ are the zeros of the partial fraction bracketed in the equation (3.1) and the multiple zeros of $f(z)$ to one lower order.*

Thus the locus of the zeros of $f'(z)$ when all the zeros of each $f_j$ vary within and on a circle $C_j$ will be composed of two parts: a part $S$ due to the zeros of the partial fraction in (3.1) and a part $T$ due to the multiple zeros of $f(z)$. The former part $S$ has already been determined. The latter part $T$ consists of every point of any region $C_j$ for which $m_j \neq 1$ and the common points of any two of the regions $C_j$, for, at such a point, two or more points $z_j$ can be made to coincide. But, since the common points of any two regions $C_j$ are already included in the part $S$ (cf. II § 5), we need not mention them again.

In short, then, *if $f_j(z)$ and $f(z) = \Pi_1^{q+1} f_j$ are polynomials of degrees $m_j$ and $n$ respectively, and if all the zeros of each $f_j$ have as common locus the interior and circumference of a circle $C_j$, the locus of the zeros of $f'(z)$ will consist of* (a) *the whole of every region $C_j$ for which $m_j \neq 1$ and* (b), *when $q \geqq 1$, the locus $R$ of* II § 5.

Reasoning based on the continuity of the zeros, similar to that used by Bôcher and Walsh, BI, p. 478 and BIV, p. 113, may now be employed to prove at once that there will exist exactly $m_j - 1$ zeros of $f'(z)$ in the region

* This is essentially Gauss' theorem, Werke, vol. 3, 1816, p. 112, that the zeros of $f'(z)$ are at the multiple zeros of $f(z)$ and at the points of equilibrium in the field of force created by unit particles at each $z_j$ repelling one another according to the inverse-distance law.

$C_j$ and $q$ in (b) if no two of the regions $C_j$ overlap and if (b) does not have any points in common with any region $C_j$; and precisely one in each part of (b) if, furthermore, (b) breaks up into $q$ distinct parts.

2. **Special cases.** As a generalization, the theorem of the preceding section should reduce for special values of $q$ and special choices of the regions $C_j$ to existing theorems on the subject. Let us see whether this is so.

When $q=0$, the locus $R$, according to the theorem, is merely the region $C_1$; that is, the zeros of $f'(z)$ lie in or on any circle enclosing the zeros of $f(z)$. This is but another way of expressing the Lucas theorem* that the zeros of $f'(z)$ lie within any convex polygon enclosing the zeros of $f(z)$.

When $q=1$, our theorem accords with one proved originally by Walsh (BII, p. 1): the locus consists of the circular regions $C_1$ (if $m_1 \neq 1$) and $C_2$ (if $m_2 \neq 1$), and of the interior and circumference of the circle

$$nm_1C_2 + nm_2C_1 - m_1m_2\sigma_{12} = 0,$$

with radius $r=(m_2r_1+m_1r_2)/(m_1+m_2)$ and center $\alpha=(m_2\alpha_1+m_1\alpha_2)/(m_1+m_2)$.

When $q=2$, our theorem again agrees with one proved by Walsh (BVI): the locus consists of the whole of every region $C_j$ for which $m_j \neq 1$ and the interior and boundary of all the ovals of the bicircular quartic (2.51).

3. **Special cases continued.** If $q$ is taken as arbitrary, but all the circles $C_j$ are given a common external center of similitude, straightforward computation† proves the curve (2.51) to break up into the circles $C_k'$ which have the same external center of similitude and the same common tangents as the circles $C_j$. The singular foci, according to their definition, are in this case the centers of the circles $C_k'$. All this agrees with a theorem due to Walsh (BV, p. 52).

Finally let us study the special case in which all the $m_j$ and all the circles $C_j$ are equal, and in which the centers of the circles $C_j$ lie at the vertices $V_j$ of a regular polygon of $q+1$ sides. Clearly the locus must be symmetric in the lines joining the center $O$ of the polygon with the points $V_j$.

If curve (2.51) does not degenerate into a set of circles with center at $O$, we may choose on curve (2.51) a point $P$ which is neither on any of the lines $OV_j$ nor on their angular bisectors. Because of the symmetry, the circle with center at $O$ and radius $OP$ would intersect the curve (2.51) in $2(q+1)$ distinct points, i.e., twice more often than permissible.

* F. Lucas, Comptes Rendus, 1868, p. 469, and 1888, p. 121.

† I have carried out the details of this computation only for the case that the external center of similitude is at the point at infinity. For $q=2$, I have also been able to show our bicircular quartic coincident with Walsh's two circles by assuming the contrary and finding that it would then intersect these two circles in a total of more than sixteen points. I hope through further study of the properties of $p$-circular $2p$-ic curves to be able to extend this geometrical method to the general case.

Hence the locus is composed of the circular regions $C_j$ if $m_j \neq 1$ and of the interior and circumference of the circle with center at $O$ and with a radius equal to the largest real root of the equation

$$\sum_0^q \frac{3}{t_j} - \sum_{j=0,k=j+1}^{q} \frac{\left\{2h \sin \frac{\pi}{q+1}(j-k)\right\}^2}{t_j t_k} = 0, \tag{3.31}$$

where

$$t_j = r^2 + h^2 - a^2 - 2hr \cos \frac{2\pi j}{q+1},$$

and where $a$ is the radius of the circle $C_j$ and $h$ that of the circumscribing circle of the polygon.

When $q=2$, equation (3.31) becomes

$$(r^2 - a^2)(r^2 + h^2 - a^2) = h^2,$$

which has as largest root $r=(a^2+ha)^{1/2}$, a number also given by Walsh (BVI, p. 140).

When $q=3$, equation (3.31) reduces to

$$(r^2 - a^2)^3 - 2h^2a^2(r^2 - a^2) - h^4a^2 = 0,$$

which, if $h=2^{1/2}$, has as largest root a number $T$,

$$(a^2 + 2a)^{1/2} < T \leqq (a^2 + sa)^{1/2}, \tag{3.32}$$

where $s>2$ and is such that

$$\frac{4}{s^3 - 4s} \leqq a. \tag{3.33}$$

Equality occurs in (3.32) when and only when it takes place in (3.33).

## IV. Special case II: Numbers $m_j$ summing to zero. Arbitrary circular regions $C_j$

**1. Preliminary remarks.** We now turn to the special case of real arbitrary $m_j$'s summing to zero.

The zeros of $F$ have in this case the striking property of being invariant under linear transformation of the plane or stereographic projection upon the sphere. If a linear transformation takes the $z_j$ into the points $z_j'$ the zeros of $F$ pass into the zeros of $F(z_j')$. If a stereographic projection throws the $z_j$ into the points $w_j$ on the sphere, the $q-1$ zeros of $F$ go into the $q-1$ positions of equilibrium on the sphere in the field of force due to particles

of mass $m_j$ at the $w_j$ repelling according to the inverse-distance law. (Cf. BI, p. 469 ff.) If, then, we project the regions comprising the locus $R$ upon the sphere, the arguments of II § 6 will serve to show that the images of these regions upon the sphere are simply-connected.

From the formal work of the second chapter, we infer at once that, when not the entire plane, the locus $R$ in case II is bounded by ovals of the curve

$$\sum_{j=1,k=j+1}^{q+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} = 0. \tag{4.1}$$

A new choice of $\Delta_{jk}$ between $\sigma_{jk}$ and $\tau_{jk}$ has however to be made in special case II for several reasons. No longer are the previous choices generally valid, because the $m_j$ and the leading coefficient of the left-hand side of equation (4.1) may now be either positive or negative, and because the regions $C_j$ and the locus $R$ may be finite or not finite.

To pass gently from chapter two, let us first make the new choice for the subcase in which both all the regions $C_j$ and the locus $R$ are finite. From this subcase we shall then proceed to the general situation by causing the regions $C_j$ of the subcase to expand continuously until they approach the $C_j$ of the general case.

But, before we may discuss this subcase, we need to obtain criteria for determining whether or not a given instance of finite regions $C_j$ will yield a finite locus $R$.

2. **Criteria for a finite $R$.** That finiteness in the regions $C_j$ does not ensure a finite locus $R$ is evident from the example

$$\frac{1}{z-1} + \frac{1}{z+1} - \frac{2}{z} \tag{4.21}$$

which has $z=\infty$ as its only zero, the regions $C_j$ being the points $z=0$ and $z=\pm 1$. (Contrast with BVIII.)

To obtain a condition for a finite $R$, let us express the equation $F=0$ in the form

$$z^{q-1}\left\{\sum_{1}^{q+1} m_j z_j\right\} + \text{terms of lower degree} = 0.$$

It is thus necessary and sufficient for $z=\infty$ to be in $R$ that there exist in each region $C_j$ a point $z_j$ making

$$\sum_{1}^{q+1} m_j z_j = 0. \tag{4.22}$$

In order to give this condition a more practical expression, let us determine the locus $L_{12}$ of the point $z_{q+1}$ defined by (4.22) when $z_1$ and $z_2$ vary respectively

over the regions $C_1$ and $C_2$. Applying the method of I § 2, we find here $R_0$ to be the null circle

$$e_0 = -\left(\sum_1^{q+1} m_j z_j\right)\left(\sum_1^{q+1} m_j \bar{z}_j\right) = \sum_{j=1,k=j+1}^{q+1} m_j m_k P_{jk} = 0,$$

and $R_2$ or $L_{12}$ finally to be the interior and circumference of the circle

$$D_{12} = m_1 m_2 \Delta_{12} + \sum_{j=1,k=3}^{j=2,k=q+1} m_j m_k C_{jk} + \sum_{j=3,k=j+1}^{q+1} m_j m_k P_{jk} = 0$$

with center at

$$-\frac{1}{m_{q+1}}\left\{\sum_1^2 m_j \alpha_j + \sum_3^q m_j z_j\right\}.$$

As the radius of $D_{12}$ is

$$\left|\frac{m_1 r_1 + m_2 r_2}{m_{q+1}}\right| \quad \text{if} \quad \Delta_{12} = \sigma_{12},$$

and

$$\left|\frac{m_1 r_1 - m_2 r_2}{m_{q+1}}\right| \quad \text{if} \quad \Delta_{12} = \tau_{12},$$

the quantity $\Delta_{12}$ must be chosen equal to $\sigma_{12}$ if $m_1 m_2 > 0$, and to $\tau_{12}$ if $m_1 m_2 < 0$.

When and only when this locus $L_{12}$ overlaps the circular region $C_{q+1}$ will condition (4.22) be fulfilled and the point $z = \infty$ belong to the locus $R$. In other words, for the locus $R$ to be finite, it is necessary and sufficient that every point $z_{q+1}$ of the region $C_{q+1}$ satisfy the inequality $D_{12}(z_{q+1}) < 0$.

3. **The finite subcase.** Let us now determine the quantity $\Delta_{jk}$ in equation (4.1) for the subcase of finite regions $C_j$ and a finite locus $R$.

From the work of the second chapter, it is plain that to determine for instance $\Delta_{12}$, it is only necessary to specialize in (4.1) all the $C_j$ except $C_1$ and $C_2$ to be null circles, and to choose between the two equations

$$E_2 = C_1 C_2 \prod_3^{q+1} P_j \left\{\frac{m_1 m_2 \sigma_{12}}{C_1 C_2} + \sum_{j=1,k=3}^{j=2,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} + \sum_{j=3,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k}\right\} = 0,$$

$$E_2' = C_1 C_2 \prod_3^{q+1} P_j \left\{\frac{m_1 m_2 \tau_{12}}{C_1 C_2} + \sum_{j=1,k=3}^{j=2,k=q+1} \frac{m_j m_k C_{jk}}{C_j P_k} + \sum_{j=3,k=j+1}^{q+1} \frac{m_j m_k P_{jk}}{P_j P_k}\right\} = 0.$$

The leading coefficient of the first is precisely $D_{12}$ with $\Delta_{12} = \sigma_{12}$ and that of the second, $D_{12}$ with $\Delta_{12} = \tau_{12}$ (cf. IV § 2).

If $m_1 m_2 > 0$ and $\zeta$ is any point on $E_2' = 0$, i.e., $E_2'(\zeta) = 0$, then

$$E_2(\zeta) = 2 m_1 m_2 [C_1 C_1]^{1/2} [C_2 C_2]^{1/2} \prod_3^{q+1} P_j(\zeta) \geqq 0. \tag{4.31}$$

Since $D_{12}<0$, inequality (4.31) signifies that all the ovals of the curve $E_2' = 0$ are contained in those of the curve $E_2 = 0$. Similarly, if $m_1m_2<0$, all the ovals of the curve $E_2=0$ are enclosed in ovals of the curve $E_2' =0$.

The choice for $\Delta_{12}$ is typical. The locus $R$ consists of the interior and boundary of all the ovals of the $(q-1)$-circular $2(q-1)$-ic curve

$$\sum_{j=1,k=j+1}^{q+1} \frac{m_jm_k\Delta_{jk}}{C_jC_k} = 0, \tag{4.32}$$

where $\Delta_{jk}=\sigma_{jk}$ if $m_jm_k>0$, and $\Delta_{jk}=\tau_{jk}$ if $m_jm_k<0$.

**4. The general case.** For convenience in our treatment of the unrestricted instance of special case II, let us introduce the notion of the "radius of a circular region $C$." Considering that direction positive along radii vectores which points into the region $C$, we define as the radius of the region $C$ ($\tilde{r}$ where $r$ is the radius of circle $C$) the algebraic distance from the circumference to the center of $C$. Thus if the region $C$ were the exterior of the circle $C$, we should find $\tilde{r}<0$.

On each circle $C_j$ let us now choose a point $z_j^*$ such that

$$\sum_{1}^{q+1} m_jz_j^* \neq 0, \tag{4.41}$$

and construct within the region $C_j$ a circle $\gamma_j$, of radius $\omega_j$, tangent to the circle $C_j$ at $z_j^*$. Because of condition (4.22) the locus of the zeros of $F$ when the $z_j$ vary over the regions $\gamma_j$ (the interior and circumference of the circle $\gamma_j$) will be finite when the radii $\omega_j$ are sufficiently small. It will be in fact bounded by curve (4.32) where the $C_j$ are replaced by the $\gamma_j$.

If we allow the $\omega_j$ gradually to increase, the coefficients of (4.32) will change continuously. If $\tilde{r}_p<0$, the stereographic projection of the circle $\gamma_p$ upon the sphere must experience two stages in order to coincide with the circle $C_p$: first the circle $\gamma_p$ must expand until it passes through the north pole of the sphere, i.e., $\tilde{\omega}_p$ must grow positively infinite; then it must continue to expand, thereafter the region $\gamma_p$ enclosing the north pole. That is to say, the quantity $\tilde{\omega}_p$ must increase from $-\infty$ to $\tilde{r}_p$.

In equation (4.32), accordingly, the forms only of

$$\sigma_{jk} = d_{jk}^2 - (r_j - r_k)^2$$

and

$$\tau_{jk} = d_{jk}^2 - (r_j + r_k)^2$$

are affected by the continuous expansions of the regions $\gamma_j$. If $\tilde{r}_j\tilde{r}_k<0$, $\sigma_{jk}$ and $\tau_{jk}$ are interchanged; otherwise, they too retain their original forms.

In other words, provided, as we have implicitly assumed, the locus $R$

is not the entire plane, it consists of a set of simply-connected regions bounded by ovals of the $(q-1)$-circular $2(q-1)$-ic curve

$$\sum_{j=1,k=j+1}^{q+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} = 0, \tag{4.42}$$

where $\Delta_{jk} = \sigma_{jk}$ if $(m_j \bar{r}_j)(m_k \bar{r}_k) > 0$ and $\Delta_{jk} = \tau_{jk}$ if $(m_j \bar{r}_j)(m_k \bar{r}_k) < 0$. In short, *the locus $R$ is always a $(q-1)$-circular domain.*

To be noticed also is the type of curve which bounds the locus $R$. Since the quantities $\Delta_{jk}$ are invariants and the $C_j(x, y)$ covariants of the forms $C_j(x_j, y_j)$, the curve (4.42) is also a covariant of these forms. This is to be expected from the remarks of IV § 1.

The locus $R$, just determined, may or may not include the point at infinity. Conditions that it should contain this point have already been found (IV § 2) in the case that all the $\bar{r}_j$ are positive. It is now desirable that we seek similar conditions for the case of unrestricted $\bar{r}_j$.

Suppose, first, that $-\infty \leqq \bar{r}_1 < 0$ but $0 < \bar{r}_2 < \infty$. By condition (4.22), when $z_1 = \infty$, $z_{q+1} = \infty$ and therefore the point $z = \infty$ will be part of the locus $L_{12}$. (Cf. IV § 2.) As $\bar{\omega}_1$ increases from $-\infty$, the quantities $\sigma_{12}$ and $\tau_{12}$ are interchanged. In any case, for large values of $-\bar{\omega}_1$, the radius of circle $D_{12}$ will be

$$-(1/|m_{q+1}|)(|m_2|\bar{\omega}_2 + |m_1|\bar{\omega}_1).$$

If, then, $\bar{\omega}_1$ continues to increase to the value $-|m_2/m_1|\bar{\omega}_2$, $D_{12}$ becomes a null circle and $L_{12}$ the entire plane. Briefly, if $r_1 \leqq |m_2/m_1| r_2$, the locus $L_{12}$ is the entire plane; if $r_1 > |m_2/m_1| r_2$, the locus $L_{12}$ is the exterior of the circle $D_{12}$ where $\Delta_{12} = \tau_{12}$ if $m_1 m_2 > 0$ and $\Delta_{12} = \sigma_{12}$ if $m_1 m_2 < 0$.

If both $\bar{r}_1 < 0$ and $\bar{r}_2 < 0$, the locus $L_{12}$ is always the entire plane. For, in allowing $\gamma_1$ and $\gamma_2$ to expand, we can choose $\bar{\omega}_1$ as negative and $\bar{\omega}_2$ as positive and so large that

$$\bar{\omega}_1 \geqq -|m_2/m_1|\bar{\omega}_2.$$

In other words, if $\bar{r}_j < 0$ for any two $j$, the locus $R$ always contains the point at infinity.*

* The point $z_{q+1}$ defined by the equation (4.22) is the center of gravity $G$ of the other points $z_j$. When the first $q$ $z_j$ vary over their regions $C_j$, the locus of $G$ will be, according to our methods,

(a) the entire plane, if, for at least two $k$, $\bar{r}_k < 0$, or if, for at least one $j$ and one $k$, $\bar{r}_j < 0$ and

$$|\bar{r}_j| \leqq |m_k/m_j|\,|\bar{r}_k|;$$

(b) the interior and circumference of the circle

$$H \equiv \sum_{\substack{i=1 \\ j=i+1}}^{q} m_i m_j \Delta_{ij} + m_{q+1} \sum_{1}^{q} m_j C_{j\,q+1} = 0$$

if $\bar{r}_k > 0$ all $k$; or,

(c) the exterior and circumference of the circle $H$ if $\bar{r}_k < 0$ for just one $k$.

The quantity $\Delta_{ij}$ is the same in $H$ as it is for the curve (4.42). For another derivation of (b), see BV, p. 60.

When $q=2$, an alternative method for deciding whether $z=\infty$ is part of the locus $R$ may be obtained directly. Equation (4.42) represents a circle which, formed for the $\gamma_j$, is

$$\Gamma = m_1m_2\Delta_{12}\gamma_3 + m_1m_3\Delta_{13}\gamma_2 + m_2m_3\Delta_{23}\gamma_1 = 0, \tag{4.421}$$

with radius and center respectively

$$\omega = \left|\frac{\nu}{\lambda}\right| \quad \text{and} \quad \alpha = \frac{\mu}{\lambda}, \tag{4.43}$$

where

$$\lambda = m_1m_2\Delta_{12} + m_1m_3\Delta_{13} + m_2m_3\Delta_{23},$$
$$\mu = m_1m_2\Delta_{12}\alpha_3 + m_1m_3\Delta_{13}\alpha_2 + m_2m_3\Delta_{23}\alpha_1,$$
$$\nu = m_1m_2\Delta_{12}\rho_3 + m_1m_3\Delta_{13}\rho_2 + m_2m_3\Delta_{23}\rho_1.$$

Since, when the $\omega_j=0$, all $j$, $\Delta_{jk}>0$, and

$$\lambda = -\left(\sum_1^3 m_j\bar{z}_j^*\right)\left(\sum_1^3 m_j\bar{z}_j^*\right) < 0,$$

we always have $\tilde{\omega} = -\nu/\lambda$. Consequently, the locus $R$ is finite or not according as $\tilde{\omega}>0$ or $\tilde{\omega}<0$.

**5. The entire plane as locus.** To complete the study of the general case, we need yet to inquire into the circumstances under which $R$ is the entire plane.

As the regions $\gamma_j$ expand, one or more of the ovals of (4.32) may conceivably expand and sweep out the entire plane; or, after $R$ has acquired the point $z=\infty$, the oval bounding the part of $R$ containing $z=\infty$ may shrink to a point and its exterior thus become the entire plane. When $q=2$, these descriptive conditions are easily expressible in terms of $\omega$ and $\alpha$ of (4.43).

The locus $R$ may also become the entire plane discontinuously. If all the regions $C_j$ have a point in common, the instant the regions $\gamma_j$ are sufficiently large all to include this common point does the equation $F=0$ become the identity

$$\left(\frac{1}{z-z_1}\right)\sum_1^{q+1} m_j = 0,$$

and therefore locus $R$ the entire plane. An example of this sort has been called to my attention by Professor Walsh. (See also B VII, p. 12.) If $q=2$, $m_1=m_2=1$, the regions $C_1$ and $C_2$ are the interior of the circle $C_1\equiv C_2$, and the region $C_3$ is merely the point at infinity, then $R$ is the locus of the zeros of

$$F = \frac{1}{z - z_1} + \frac{1}{z - z_2}$$

or the locus of the point $z = \frac{1}{2}(z_1 + z_2)$. Obviously $R$ is the interior of the circle $C = C_1$. If the region $C_1$ is suitably allowed to expand, the region $C$ remains coincident with it, until the region $C_1$ becomes a half-plane when, because all three regions $C_j$ have a common point, the region $C$ suddenly becomes the entire plane.

## V. Applications

**1. A cross-ratio theorem.** We now return to the theorem which was the starting point of this paper (cf. I §1), the theorem that "if the points $z_1$, $z_2$, $z_3$ vary independently and have circular regions as their respective loci, then the locus of the point $z$ defined by the equation

$$\frac{m_1}{z - z_1} + \frac{m_2}{z - z_2} + \frac{m_3}{z - z_3} = 0, \tag{5.1}$$

where $m_1$, $m_2$, $m_3$ are real constants whose sum is zero, is also a circular region."

By specializing throughout the preceding chapter the number $q$ to be two, we obtain at once (1) an analytical proof of this theorem; (2) conditions that the locus contain the point at infinity; (3) conditions that the locus be the entire plane; and (4) the equation of the circle bounding the locus when the locus is not the entire plane. The latter equation is (cf. equation (4.42))

$$m_1 m_2 \Delta_{12} C_3 + m_1 m_3 \Delta_{13} C_2 + m_2 m_3 \Delta_{23} C_1 = 0,$$

where $\Delta_{jk} = \sigma_{jk}$ if $(m_j \bar{r}_j)(m_k \bar{r}_k) > 0$ and $\Delta_{jk} = \tau_{jk}$ if $(m_j \bar{r}_j)(m_k \bar{r}_k) < 0$.

As we remarked in the fifth footnote to chapter I, the above theorem may also be expressed in terms of a real constant cross ratio, instead of in terms of the partial fraction (5.1).

**2. The jacobian of two binary forms.** A second immediate application of the results of the preceding chapter is to the location of the zeros of the jacobian of the two binary forms

$$f_1 = \prod_1^{p_1} (uv_j - vu_j) \quad \text{and} \quad f_2 = \prod_{p_1+1}^{p_1+p_2} (uv_j - vu_j).$$

By Euler's law for homogeneous functions, this jacobian $J_{12}$ may be written as

$$J_{12} = \frac{f_1 f_2}{v^2} \left\{ \sum_1^{p_1} \frac{p_2 v_j v}{uv_j - vu_j} + \sum_{p_1+1}^{p_1+p_2} \frac{(-p_1) v_j v}{uv_j - vu_j} \right\}. \tag{5.21}$$

The zeros of $J_{12}$ are thus either multiple zeros of one lower order of the pro-

duct $f_1 f_2$, or zeros of the bracket in (5.21). (Cf. BI, p. 476.) On introduction of the non-homogeneous coördinates $z = u/v$, this bracket becomes the partial fraction

$$F = \sum_{1}^{p_1} \frac{p_2}{z - z_j} + \sum_{p_1+1}^{p_1+p_2} \frac{(-p_1)}{z - z_j}.$$

As in III §1, we now obtain the following theorem:

*If $f_{ij}$, $f_1 = \prod_1^i f_{1j}$, and $f_2 = \prod_{i+1}^{q+1} f_{2j}$ are binary forms of degrees $n_{ij}$, $1/p_1, -1/p_2$ respectively, and if all the zeros of each $f_{ij}$ have as common locus a circular region $C_j$, the locus of the zeros of $J_{12}$ will consist of* (a) *the whole of every region $C_j$ for which $n_{ij} \neq 1$, and* (b), *when $q \geqq 2$, the $(q-1)$-circular domain specified in* IV §4 *and* §5, *where $m_j = p_i n_{ij}$.*

Reasoning similar to that referred to in III §1 will also establish the facts that there will exist in each region $C_j$ exactly $n_{ij} - 1$ zeros of $J_{12}$ and $q-1$ in (b) if no two of the regions $C_j$ have any point in common and if, in addition, (b) does not overlap any region $C_j$; and precisely one zero in each part of (b) if furthermore (b) falls into $q-1$ distinct parts.

When $q = 1$, the theorems of this section become identical with those of Bôcher (BI, p. 477) that, if all the zeros of $f_1$ and $f_2$ lie respectively in the circular regions $C_1$ and $C_2$, there are no zeros of $J_{12}$ outside of these two regions, and that, if the regions $C_1$ and $C_2$ do not overlap, there are exactly $p_1 - 1$ zeros of $J_{12}$ in the first and $p_2 - 1$ zeros of $J_{12}$ in the second region.

When $q = 2$, the theorems of this section reduce to those due to Walsh (BIV, p. 112) that, if the circular regions $C_1$, $C_2$, $C_3$ are the respective loci of $m_1$ roots of $f_1$, the remaining $m_2$ roots of $f_1$, and all the $m_3$ roots of $f_2$, then the locus of zeros of $J_{12}$ will consist of the whole of each circular region $C_j$ for which $m_j \neq 1$ and the locus described in V §1.

Similar theorems hold true for the covariant

$$\Phi = \frac{1}{k}\left(\prod_1^k f_j\right) \sum_{i,j=1}^{k} \frac{m_i}{p_i} \frac{J_{ij}}{f_i f_j}$$

of the forms

$$f_j = \prod_{P_{j-1}+1}^{P_j} (uv_i - vu_i)$$

where $P_0 = 0$, and $P_j - P_{j-1} = p_j$, the degree of $f_j$, and $\sum_1^k m_i p_i = 0$. For, we may write

$$\Phi = \frac{1}{v^2}\left(\prod_1^k f_j\right) \sum_1^k m_j \sum_{P_{j-1}+1}^{P_j} \frac{vv_j}{uv_j - vu_j}.$$

(See BI, p. 467.)

## VI. The most general case

**1. The locus $R$.** Having obtained the locus $R$ in the special case II (the case of real numbers $m_j$ summing to zero but of arbitrary circular regions $C_j$) we are now in a position to study the locus $R$ in the case of unrestricted numbers $m_j$ and of unrestricted circular regions $C_j$. For, as we shall presently see, the results in the most general case can be deduced from those of special case II.

Let us consider the partial fraction

$$F = \sum_{1}^{q+1} \frac{m_j}{z - z_j} + \frac{m_{q+2}}{z - z_{q+2}}, \tag{6.11}$$

where $n+m_{q+1}=0$ and $n=\sum_1^{q+1} m_j$. This fraction is of the type considered in the special case II, but, if the point $z_{q+2}$ is required to recede to infinity, the term in the subscript $q+2$ disappears from expression (6.11) and the fraction becomes one in which the sum of the $m_j$'s is not necessarily zero.

In the particular instance of case II in which $q+2$ circular regions $C_j$ appear and the region $C_{q+2}$ consists merely of the point $z_{q+2}$, the locus $R$ is bounded by the curve (cf. equation (4.42))

$$\sum_{j=1,k=j+1}^{q+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} + \sum_{1}^{q+1} \frac{m_{q+2} m_j C_{j\,q+2}}{P_{q+2} C_j} = 0. \tag{6.12}$$

Let us now cause the point $z_{q+2}$ to move off to infinity. The above-considered locus $R$ will then become the locus $R$ of the unrestricted case. Since

$$\frac{C_{j\,q+2}}{P_{q+2}} = \frac{x_{q+2}^2 + y_{q+2}^2 - 2a_j x_{q+2} - 2b_j y_{q+2} + c_j}{x_{q+2}^2 + y_{q+2}^2 - 2x x_{q+2} - 2y y_{q+2} + x^2 + y^2}$$

and

$$\lim_{|z_{q+2}|\to\infty} \left(\frac{C_{j\,q+2}}{P_{q+2}}\right) = 1,$$

the locus $R$ of the general case will therefore be bounded by the curve

$$\sum_{j=1,k=j+1}^{q+1} \frac{m_j m_k \Delta_{jk}}{C_j C_k} - \sum_{1}^{q+1} \frac{n m_j}{C_j} = 0, \tag{6.13}$$

where $\Delta_{jk}=\sigma_{jk}$ if $(m_j\tilde{r}_j)(m_k\tilde{r}_k)>0$ and $\Delta_{jk}=\tau_{jk}$ if $(m_k\tilde{r}_k)(m_k\tilde{r}_k)<0$.

Thus we obtain a curve which clearly includes as special cases the curves (2.51) and (4.42) and which has, if $n\neq 0$, the same singular foci as curve (2.51).

In the same manner, we can carry over from the special case II all of

the other important properties of the locus $R$. We see thereby that *in every case the locus $R$ is a $p$-circular domain, where*

$$p = q - 1 + \lim_{\nu=0} \left( \frac{n}{n+\nu} \right).$$

**2. Application to the derivative of a rational function.** As a final application let us consider the location of the zeros of the derivative of the rational function

$$f(z) = \frac{\prod_{1}^{u}(z - z_j)}{\prod_{u+1}^{u+v}(z - z_j)},$$

where the numerator $z_j$ are distinct from the denominator ones. Since

$$f'(z) = f(z)\left\{ \sum_{1}^{u} \frac{1}{z - z_j} - \sum_{u+1}^{u+v} \frac{1}{z - z_j} \right\}, \tag{6.21}$$

the zeros of $f'(z)$ are the multiple zeros of $f(z)$ to one lower order, the point at infinity if $v-u>0$ and the zeros of the partial fraction in the brace. We may accordingly state the following theorem, which, when $q=1$ or 2, reduces to those proved by Walsh in BIV, pp. 114–115:

*If $f_{1j}(z)$, $f_{2i}(z)$, $f_1=\Pi_1^s f_{1j}$, and $f_2=\Pi_{s+1}^{q+1} f_{2i}$ are polynomials of degrees $m_j$, $-m_i$, $u$ and $v$ respectively; if $f_1$ and $f_2$ have no zeros in common, and if all the zeros of each polynomial $f_{ik}$ have as common locus a circular region $C_k$, then the locus of the zeros of the derivative of the rational function $f(z)=f_1/f_2$ will be*

(a) *the whole of each region $C_j$ for which $m_j \neq 1$ and $j=1, 2, 3, \cdots, s$;*

(b) *the point at infinity if $v-u>0$, and*

(c) *if $q \geqq 1$, the $p$-circular domain specified in the preceding section*, VI § 1.

## VII. Generalizations

**1. Complex $m_j$.** The methods of the preceding chapters permit us also to study the locus $R$ of the zeros of the partial fraction $F$ when the numbers $m_j$ are complex.* We give here only the results in two special cases: case (a), where $q=1$, $m_1+m_2 \neq 0$, and the circular regions $C_j$ are finite; and case (b), where $q=2$, $m_1+m_2+m_3=0$, and both the regions $C_j$ and the locus $R$ are finite.

* Cf. BVII for a geometrical discussion of cases of complex $m_j$, and BIX for a study of simple zero-free regions of the fraction $F$ in the case of complex $m_j$; also Walsh, these Transactions, vol. 24 (1922), pp. 163–180; p. 169.

We shall in the statement of these results use the following additional notation:

$$\mu_{jk} = \frac{1}{2}(\bar{m}_j m_k + \bar{m}_k m_j)\ ; \quad \nu_{jk} = \frac{i}{2}(\bar{m}_j m_k - m_j \bar{m}_k)\ ;$$

$$l_{z_j z_k} = i(\bar{z}_j z_k - \bar{z}_k z_j)\ ;$$

$$L_{jk} = i\{(\bar{z} - \bar{\alpha}_j)(z - \alpha_k) - (z - \alpha_j)(\bar{z} - \bar{\alpha}_k)\}.$$

In subcase (a), the locus $R$ is the interior and circumference of the circle

$$(|m_1|^2 + \mu_{12})C_2 + (|m_2|^2 + \mu_{12})C_1 - \{\mu_{12}\sigma_{12} + 2(|m_1 m_2| - \mu_{12})r_1 r_2 - \nu_{12}L_{12}\} = 0,$$

with center at the point

$$\frac{1}{|m_1 + m_2|^2}\{(|m_1|^2 + \mu_{12})\alpha_2 + (|m_2|^2 + \mu_{12})\alpha_1\}$$

and with a radius equal to

$$\frac{|m_1| r_2 + |m_2| r_1}{|m_1 + m_2|}.$$

In subcase (b), the subcase of V § 1 for complex $\lambda$, the locus $R$ will be finite provided that the region $C_3$ does not overlap the interior or circumference of the circle

$$E = \mu_{23}C_{23} + \mu_{13}C_{13} + \mu_{12}\tau_{12} + 2(|m_1|\,|m_2| + \mu_{12})r_1 r_2 - \nu_{23}(l_{\alpha_1\alpha_2} + l_{\alpha_2\alpha_3} + l_{\alpha_3\alpha_1}) = 0,$$

i.e., provided that, for all points $z_3$ in the region $C_3$,

$$E(z_3) < 0.$$

The locus $R$ then will consist of the interior and circumference of the circle

$$H = \{\mu_{12}\tau_{12} + 2(|m_1|\,|m_2| + \mu_{12})r_1 r_2 + \nu_{12}L_{12}\}C_3 + \{\mu_{31}\tau_{31} + 2(|m_3|\,|m_1| + \mu_{31})r_3 r_1 + \nu_{31}L_{31}\}C_2 + \{\mu_{23}\tau_{23} + 2(|m_2|\,|m_3| + \mu_{23})r_2 r_3 + \nu_{23}L_{23}\}C_3 = 0.$$

In this equation the linear expressions $L_{jk}$ do not prevent $H=0$ from representing a circle.

As is easily seen, the results for both subcases agree with our previous ones if the $m_j$ are made real numbers.

2. **Zeros of $F-\lambda$.** Likewise our methods permit us to secure the locus of the zeros of $F-\lambda$, where $\lambda$ is an arbitrary constant, real or complex. As

a sample of such results, we state the following theorem for the case of real positive $m_j$ and of finite regions $C_j$.

*The locus of the zeros of $F-\lambda$ is the interior and the boundary of all the ovals of the $(q+1)$-circular $2(q+1)$-ic curve*

$$\lambda\bar{\lambda} - \sum_1^{q+1} \frac{m_j L_j}{C_j} + \sum_1^{q+1} \frac{n m_j}{C_j} - \sum_{j=1, k=j+1}^{q+1} \frac{m_j m_k \sigma_{jk}}{C_j C_k} = 0, \tag{7.21}$$

*where $n=\sum_1^{q+1} m_j$, and*

$$L_j = \bar{\lambda}(\bar{z} - \bar{\alpha}_j) + \lambda(z - \alpha_j).$$

*The singular foci of* (6.21) *are the zeros of $F-\lambda$ for $z_j=\alpha_j$.*

As immediate applications of these results, we mention the location of the zeros of linear combinations of either a rational function and its derivative or the product of two binary forms and the jacobian of these forms.* For instance, the expression $f'(z)-\lambda f(z)$, where $f(z)$ is a rational function and $\lambda$ an arbitrary constant, has its zeros at the multiple roots of $f(z)$ and at the zeros of $F-\lambda$, where $F$ is the partial fraction $f'/f$.

In subsequent papers we hope to study further the properties of the locus $R$ and to give additional specializations and generalizations.

---

* Cf. Walsh, Bulletin of the American Mathematical Society, vol. 30 (1924), pp. 51–62, for a study of the zeros of linear combinations of a polynomial and $k$ of its derivatives.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# ON DIFFERENTIAL EQUATIONS ADMITTING POLYGENIC INTEGRALS*

BY

GEORGES CALUGARÉANO

We have shown, in a previous paper,† that some differential equations admit more general integrals than those given by the theorem of Cauchy; indeed, this theorem which uses majorant functions gives analytic integrals for each differential equation and we have seen that there are equations which admit polygenic integrals. But in the paper mentioned above we did not succeed in defining completely the class to which those equations belong and which we shall denote by class $(\alpha)$. We shall show that this class may be completely defined by a property of its general monogenic integral.

1. Let us consider the second-order equation.

$$y'' = F(z, y, y') \tag{1}$$

which belongs to the class $(\alpha)$, admitting the polygenic integral

$$y = \phi(z, \bar{z}), \tag{2}$$

where $\phi$ is a monogenic function in respect to each of the variables $z$ and $\bar{z}$. If $\mu = e^{-2i\theta}$ we have

$$y' = \phi_z + \mu\phi_{\bar{z}} \; ; \; y'' = \phi_{z^2} + 2\mu\phi_{z\bar{z}} + \mu^2\phi_{\bar{z}^2}, \tag{3}$$

and eliminating $\mu$ we get

$$y = \phi_{z^2} + 2\phi_{z\bar{z}}\frac{y' - \phi_z}{\phi_{\bar{z}}} + \phi_{\bar{z}^2}\left[\frac{y' - \phi_z}{\phi_{\bar{z}}}\right]^2. \tag{4}$$

Substituting in (4) the value of $\bar{z}$ deduced from (2) we obtain an equation of the following form:

$$y'' = A(z, y)\cdot y'^2 + B(z, y)\cdot y' + C(z, y), \tag{5}$$

hence

$$F(z, y, y') = A(z, y)\cdot y'^2 + B(z, y)\cdot y' + C(z, y), \tag{5'}$$

which holds good for any value of $\theta$, hence for any value of $y'$ on the corre-

* Presented to the Society, December 27, 1929; received by the editors in September, 1929.

† See these Transactions, vol. 31, p. 372.

sponding Kasner* circle. The expression (5′) is therefore an identity in $y'$, and hence any equation of the second order belonging to class $(\alpha)$ is of form (5). We get then easily the supplementary conditions for the determination of $A, B, C$.

It is to be remarked that the elimination made above holds without any modification if we start from

$$y = \phi(z, \gamma_0 \bar{z} + \gamma_1 z + \gamma_2), \tag{6}$$

where $\gamma_0$, $\gamma_1$ and $\gamma_2$ are arbitrary constants.

Indeed, if $k = \gamma_0 \bar{z} + \gamma_1 z + \gamma_2$ we have

$$y' = \frac{\partial \phi}{\partial z} + (\gamma_1 + \mu\gamma_0)\frac{\partial \phi}{\partial k}\,; \cdots;\; y^{(n)} = \left(\frac{\partial}{\partial z} + (\gamma_1 + \mu\gamma_0)\frac{\partial}{\partial k}\right)^{(n)} \phi, \tag{7}$$

and the elimination of $\gamma_1 + \mu\gamma_0$ and $k$ between (6) and (7) is equivalent to the elimination of $\mu$ and $\bar{z}$ between (2) and (3). But (6) shows that $y$ has a special property. We know that the monogenic integral of (1) is obtained from (6) by replacing $\bar{z}$ by $z - C$. The same result is obtained if we make $\gamma_0 = 0$, and this shows that the monogenic integral of (1) is of the form

$$Y = \phi(z, \gamma_1 z + \gamma_2). \tag{8}$$

This integral depends in a special manner on the arbitrary constants $\gamma_1$ and $\gamma_2$. This property furnishes the characterization of the whole class $(\alpha)$.

2. Let

$$y = \phi(z, \bar{z}, \gamma_3, \gamma_4, \cdots, \gamma_n) \tag{9}$$

be a polygenic function depending on $n-2$ arbitrary constants, $\gamma_3, \gamma_4, \cdots, \gamma_n$. We can write

$$y' = \frac{\partial \phi}{\partial z} + \mu\frac{\partial \phi}{\partial \bar{z}}\,; \cdots; y^{(n)} = \left(\frac{\partial}{\partial z} + \mu\frac{\partial}{\partial \bar{z}}\right)^{(n)} \phi, \tag{10}$$

and the elimination of the parameters $\mu, \bar{z}, \gamma_3, \cdots, \gamma_n$ between the equations (9) and (10) will give the most general equation of order $n$ belonging to the class $(\alpha)$. But the preceding elimination can be performed without any modification if we start from

$$y = \phi(z, \gamma_0 \bar{z} + \gamma_1 z + \gamma_2, \gamma_3, \gamma_4, \cdots, \gamma_n), \tag{11}$$

$$y' = \frac{\partial \phi}{\partial z} + \mu_1\frac{\partial \phi}{\partial k}\,; \cdots; y^{(n)} = \left(\frac{\partial}{\partial z} + \mu_1\frac{\partial}{\partial k}\right)^{(n)} \phi, \tag{12}$$

where $k = \gamma_0 \bar{z} + \gamma_1 z + \gamma_2$ and $\mu_1 = \gamma_1 + \gamma_0 \mu$.

* See E. Kasner, Science, vol. 66 (1927), pp. 581–582; Proceedings of the National Academy of Sciences, vol. 13 (1928), pp. 75–82; E. R. Hedrick, ibid., vol. 14.

Let us denote by $E_1$ all the above eliminations. If $\gamma_0$ is replaced by zero in (11), and if we form the differential equation whose general integral is the monogenic function thus obtained, we get

$$Y = \phi(z, \gamma_1 z + \gamma_2, \gamma_3, \cdots, \gamma_n), \tag{13}$$

$$\frac{dY}{dz} = \frac{\partial \phi}{\partial z} + \gamma_1 \frac{\partial \phi}{\partial k_1},$$

. . . . . (14)

$$\frac{d^n Y}{dz^n} = \left(\frac{\partial}{\partial z} + \gamma_1 \frac{\partial}{\partial k_1}\right)^{(n)} \phi,$$

where $k_1 = \gamma_1 z + \gamma_2$.

It is obvious that the elimination of $\gamma_1, k_1, \gamma_3, \gamma_4, \cdots, \gamma_n$ between (13) and (14), which elimination we will call $E_2$, will give the same result as $E_1$. Therefore any equation of the $n$th order belonging to the class $(\alpha)$ has a monogenic integral of the form

$$Y(z) = \phi(z, \gamma_1 z + \gamma_2, \gamma_3, \cdots, \gamma_n). \tag{15}$$

The converse is also true; if a differential equation has this property, it belongs to the class $(\alpha)$. Indeed, if (15) is its monogenic integral, let us form (11) by adding $\gamma_0 \bar{z}$ to $\gamma_1 z + \gamma_2$. The eliminations $E_1$ and $E_2$ giving always the same result, it follows that (11) is the polygenic integral of the same equation.

Hence, *the property which determines the class* $(\alpha)$ *is that its monogenic integral contains in a special way two arbitrary constants.*

It is obvious that (11) gives the most general polygenic integral of the equation of the $n$th order belonging to $(\alpha)$. Because, if $y(z)$ contains more than $n+1$ distinct arbitrary constants, their elimination could not give an equation of the $n$th order. Therefore, (11) is the form of the general polygenic integral of the equation of the $n$th order of $(\alpha)$. It contains $n+1$ arbitrary constants; the solution of Cauchy is a particular case obtained for $\gamma_0 = 0$. Inversely, if an equation of the class $(\alpha)$ is given, and if its monogenic general integral is known, the polygenic integral is obtained by adding the term $\gamma_0 \bar{z}$ to $\gamma_1 z + \gamma_2$.

3. In the preceding theory we have used *rectilinear* derivations; that is, we made $dz_\theta$ tend to zero in the expression for the higher derivatives

$$\frac{d^n y}{dz_\theta^n} = y^{(n)} = \left(\frac{\partial}{\partial z} + e^{-2i\theta} \frac{\partial}{\partial \bar{z}}\right)^{(n)} y,$$

the affix of $dz_\theta$ lying on the straight line which passes through $z$ and makes the

angle $\theta$ with $Ox$. But we can generalize this conception by considering, as did Kasner,* the "curvilinear" derivatives where the path of approach is any curve. It may be interesting to see which are the classes of differential equations which admit polygenic integrals, when we take the derivatives of $y$ in the general conception of Kasner, taking into account the curvature $\kappa$, and even other differential elements of a higher order, of the path curve on which $z+dz_\theta$ tends to $z$, this path curve being considered as entirely arbitrary.

It is obvious that the types thus obtained all belong to class $(\alpha)$.

---

* See his paper *The second derivative of a polygenic function*, these Transactions, vol. 30 (1928), pp. 803–818.

UNIVERSITY OF BUCHAREST,
BUCHAREST, ROUMANIA

# DIVERGENT SERIES AND SINGULAR POINTS OF ORDINARY DIFFERENTIAL EQUATIONS*

BY

GEORGE D. BIRKHOFF AND FREDERIC R. BAMFORTH†

## INTRODUCTION

The analytic form of the solutions of the system of differential equations

$$\frac{dx_1}{X_1(x_1, \cdots, x_n)} = \cdots = \frac{dx_n}{X_n(x_1, \cdots, x_n)} \quad (1)$$

in a neighborhood of the point $x_1 = \cdots = x_n = 0$, in which all the functions $X_i$ are supposed to be analytic, $X_i(0, \cdots, 0)$ being zero for $i = 1, 2, \cdots, n$, has been the subject of much study. This is justified by the applications which can be made of this form to various theories in analysis and in dynamics. Dulac‡ has simplified the problem in many cases by reducing the equations (1) to simple reduced forms of which the integration can be made without difficulty. The integration of these reduced equations then furnishes the solution of the system (1) either in terms of a parameter or in the form of a system of integrals.

Let $m_i$ be the roots of the so-called characteristic equation which, when written in determinant form, is

$$\left| \frac{\partial X_i}{\partial x_j} - m\delta_{ij} \right|_{x_1 = \cdots = x_n = 0} = 0,$$

where $\delta_{ij} = 0$ if $i \neq j$ and 1 if $i = j$, and let $L_i$ represent the linear terms of $X_i$, $i = 1, 2, \cdots, n$, when these functions are expanded according to powers of $x_1, \cdots, x_n$. Three conditions which have played an important rôle in the study of the system (1) may now be written as follows: (i) if the numbers $m_i$ are represented on a complex plane, there exists a straight line passing through the origin which is such that all the points $m_i$ are on the same side of it; (ii) there exists a linear change of variables $x_j = x_j(y_1, \cdots, y_n)$ such that the system of differential equations

$$\frac{dx_i}{d\tau} = L_i$$

* Presented to the Society, March 29, 1929; received by the editors May 15, 1929.

† National Research Fellow.

‡ Two papers to which reference will be made in the text are the following: I: H. Dulac, Bulletin de la Société Mathématique de France, vol. 40 (1912), p. 324 et seq.; II: G. D. Birkhoff, Berlin Sitzungsberichte, 1929, pp. 171–183.

is equivalent to the system

$$\frac{dy_i}{d\tau} = m_i y_i ;$$

(iii) there exists among the numbers $m_i$ no relationship of the form

$$m_i = p_1 m_1 + \cdots + p_n m_n, \tag{2}$$

$p_j$ being positive integers whose sum is greater than 1. We shall refer to these conditions frequently.

On introducing a parameter $t$ it is seen that the system (1) can be reduced to the system

$$\frac{dx_i}{dt} = X_i(x_1, \cdots, x_n) \qquad (i = 1,2, \cdots, n). \tag{3}$$

The following lemma can be readily proved* by the aid of a theorem of Poincaré's and is already in the literature:

LEMMA A. *When the conditions* (i), (ii) *and* (iii) *are fulfilled, the system of differential equations* (3) *is equivalent in a neighborhood of* $x_1 = \cdots = x_n = 0$ *to the system*

$$\frac{dz_i}{dt} = m_i z_i \qquad (i = 1,2, \cdots, n), \tag{4}$$

*by means of a one-to-one analytic transformation of the form*

$$z_i = \phi_i(x_1, \cdots, x_n) \qquad (i = 1,2, \cdots, n), \tag{5}$$

*which leaves the point* $x_1 = x_2 = \cdots = x_n = 0$ *invariant.*

That this transformation is analytic is due to the fact that the formal power series solutions of the partial differential equations

$$\frac{\partial \phi_i}{\partial x_1} X_1 + \cdots + \frac{\partial \phi_i}{\partial x_n} X_n = m_i \phi_i \qquad (i = 1, \cdots, n) \tag{6}$$

converge for $|x_i| < r$, $r > 0$. But when the $m_i$ satisfy the conditions (ii) and (iii) and not the condition (i) in the case of real variables, the power series solutions of the equations (6) may diverge. Nevertheless, as has been conjectured,† even though these series diverge there may exist transformations playing the same rôle in which the functions $\phi_i$ are of class $C^\infty$ in a neighborhood of the origin and are analytic at every point of this neighborhood except

* I, p. 328.

† See II.

possibly those which lie on certain manifolds of dimensions less than $n$. The existence of such a transformation has been proved* for the outstanding case $n=2$ in which $m_1$ is greater than zero, and $m_2$ is less than zero. In the present paper an analogous transformation is obtained for the case $n=3$ in which the $m_i$ are real and not all of one sign. It seems likely that the remaining important case $n=3$ not yet treated, in which $m_2$ and $m_3$ are conjugate imaginaries with real part of opposite sign to $m_1$, will offer essentially no new difficulties.

The case $n=3$ here treated differs essentially from the case $n=2$ in that for the former it is necessary to define two so-called invariant surfaces. The cases for $n>3$ seem to offer further difficulties. These will be discussed more fully in another paper.

Since in Section I we shall make use of another lemma which is already in the literature,† we shall quote it here in a restricted form as

LEMMA B. *When the right hand sides of the differential equations* (3) *have the form* $m_ix_i+X_{2i}$, $i=1, 2, \cdots, n$, *where* $X_{2i}$ *have no terms of degree less than two,* $m_1, \cdots, m_p$ *are greater than zero,* $m_{p+1}, \cdots, m_n$ *are less than zero, and the condition* (iii) *is satisfied, there exists a transformation of the form*

$$y_i = x_i + f_i(x_{p+1}, \cdots, x_n) \qquad (i = 1, \cdots, p),$$
$$y_j = x_j \qquad (j = p+1, \cdots, n),$$

*in which the functions* $f_i$ *are analytic in their variables in a neighborhood of* $x_{p+1}= \cdots =x_n=0$, *which reduces the differential equations* (3) *to ones of the same type in* $y_k$ *with the additional property that the function on the right hand side of any* $i$*th equation vanishes when* $y_1= \cdots =y_p=0$.

## I. PRELIMINARY ANALYTIC NORMALIZATION

Consider the system of differential equations

$$\frac{dx_i}{dt} = X_i(x_1, x_2, x_3) \qquad (i = 1,2,3), \tag{7}$$

in which the $X_i$ are real power series in the $x_j$, have no constant terms, and converge for $|x_j| \leqq r_1$ where $r_1>0$. The coefficients in the $X_i$ are constants and not functions of the parameter $t$. Let $m_1$, $m_2$ and $m_3$ be the roots of the determinant equation

$$\left| \frac{\partial X_i}{\partial x_j} - m\delta_{ij} \right|_{x_1=x_2=x_3=0} = 0.$$

* Loc. cit. II.

† I, p. 359 et seq.

It will be assumed that $m_1$ and $m_2$ are greater than zero, $m_3$ is less than zero, and that the conditions (ii) and (iii) mentioned in the Introduction are satisfied.

The object of this paper is to show that the system of differential equations (7) is equivalent to the system

$$\frac{dz_i^*}{dt} = m_i z_i^* \qquad (i = 1,2,3), \tag{8}$$

by means of a one-to-one transformation of the type

$$z_i^* = \phi_i(x_1, x_2, x_3) \tag{9}$$

where the functions on the right involve the displayed arguments only, are zero at the origin, $x_1=x_2=x_3=0$, are continuous together with all their partial derivatives within a certain neighborhood of the origin, are analytic at any point of this neighborhood which does not lie on any one of a certain set of three surfaces which pass through the origin, and have certain further analyticity properties. If $m_3$ were positive instead of negative and the conditions (ii) and (iii) were still satisfied, then the transformation (9) might be chosen analytic in a neighborhood of the origin as Lemma A shows. However, when $m_3$ is negative, a transformation of the type just mentioned is the closest approach to an analytic transformation which we have discovered that will reduce the equations (7) to the equations (8). The equivalence of these two systems will only be shown to exist for a certain small neighborhood of the origin.

By an "invariant surface" will be meant a surface in the space of the dependent variables which is made up of integral curves of the differential equations (7) and contains the origin as an interior point; by an "invariant curve" will be meant an integral curve of (7) which passes through the origin. Now the classical theory of the system of differential equations (7) tells us that there exists an analytic invariant surface which contains two analytic invariant curves, and that there exists a third analytic invariant curve which does not lie in this surface. The first sequence of transformations which will be used will reduce the equations (7) to a special form in which these three analytic invariant curves are the three axes in the space of the dependent variables and the analytic invariant surface is one of the coordinate planes.

On account of the assumption that the condition (ii) is satisfied, we may assume that the right hand members of the differential equations (7) have the forms

$$X_i = m_i x_i + X_{2i}(x_1, x_2, x_3) \qquad (i = 1,2,3), \tag{10}$$

in which the $X_{2i}$ have no terms of less than the second degree in their arguments when they are expanded as power series about the origin.

Now let us examine the differential equations (7) whose right hand members have the forms given in (10) in the light of Lemmas A and B of the Introduction. Since we may perform the two analytic transformations which Lemma B assures us exist, the first being obtained when we consider $m_1$, $m_2$ as the $m_1, \cdots, m_p$ of the lemma, and the second when we change $t$ to $-t$ and consider $m_3$ as the $m_1, \cdots, m_p$ of the lemma, we may assume that the $X_{2i}$ of (10) have the property that

$$X_{21}(0,0,x_3) = X_{22}(0,0,x_3) = X_{23}(x_1,x_2,0) = 0. \tag{11}$$

The performing of these transformations moves the third analytic invariant curve and the analytic invariant surface, which have been mentioned before, into the $x_1=x_2=0$ axis and the $x_3=0$ plane, respectively.

We wish now to move the two analytic invariant curves in the $x_3=0$ plane into the coördinate axes of that plane as well as make further simplifications. Since we wish to work in the $x_3=0$ plane, let us set $x_3=0$ which is seen to satisfy the third differential equation of (7) on account of the equations (11). The other two equations of (7) then take the form

$$\frac{dx_1}{dt} = m_1x_1 + X_{21}(x_1,x_2,0), \quad \frac{dx_2}{dt} = m_2x_2 + X_{22}(x_1,x_2,0).$$

In as much as these equations satisfy the hypotheses of Lemma A, and we may perform the transformation which that lemma says exists, we may assume that

$$X_{21}(x_1,x_2,0) = X_{22}(x_1,x_2,0) = 0$$

which implies that the analytic invariant curves in the $x_3=0$ plane are the coördinate axes.

Since, on account of the equations (11), which are still true after the transformation just mentioned has been performed, $x_1=0$, $x_2=0$ satisfies the first two equations of (7), we may assume that $X_{23}(0, 0, x_3)=0$, because if it were not so we could make use of an analytic transformation of the form

$$y_1 = x_1, \quad y_2 = x_2, \quad y_3 = x_3 + f(x_3),$$

whose existence is assured by Lemma A, to make $Y_{23}(0, 0, y_3)=0$.

As a result of this discussion we may state that *without loss of generality, the differential equations* (7) *may be assumed to have the form*

$$\frac{dx_1}{dt} = x_1(m_1 + x_3P_1) + x_2x_3Q_1,$$

$$\frac{dx_2}{dt} = x_2(m_2 + x_3P_2) + x_1x_3Q_2, \tag{12}$$

$$\frac{dx_3}{dt} = x_3(m_3 + x_1P_3 + x_2P_4),$$

*where the $P_i$ and $Q_i$ are analytic functions of their arguments $x_j$ in a neighborhood of the origin, for $|x_i| \leqq r$, $r>0$, $i=1, 2, 3$, $Q_1$ is a function of $x_2$ and $x_3$ only, and $Q_2$ is a function of $x_1$ and $x_3$ only.*

## II. Convergence properties of formal series

Let us recall that the aim of the present paper is to find three functions $\phi_i(x_1, x_2, x_3)$ such that, when the $x_j$ are replaced by an arbitrary solution $x_j(t)$ of the differential equations (12), $z_i^* = \phi_i[x_1(t), x_2(t), x_3(t)]$ satisfy the differential equations (8). Since all three of these functions can be found in exactly the same way we shall confine our attention for the present to $\phi_1$. Evidently $\phi_1$ must be a solution of the partial differential equation

$$\frac{\partial\phi_1}{\partial x_1}[x_1(m_1 + x_3P_1) + x_2x_3Q_1] + \frac{\partial\phi_1}{\partial x_2}[x_2(m_2 + x_3P_2) + x_1x_3Q_2] + \frac{\partial\phi_1}{\partial x_3}[x_3(m_3 + x_1P_3 + x_2P_4)] = m_1\phi_1. \tag{13}$$

Equation (13) has a formal power series solution whose only linear term is $ax_1$, where $a$ is an arbitrary constant. This power series solution is unique if the coefficient of $x_1$ is chosen as 1. Let us assume that this has been done. Furthermore, every other term has a factor $x_1x_3$ or $x_2x_3$ as is readily seen by alternately setting $x_1=x_2=0$ and $x_3=0$ in (13) and solving for $\phi_1$. Let this formal series be

$$\phi_1 \sim x_1 + \cdots + \frac{a_{mnp}}{m!n!p!}x_1^m x_2^n x_3^p + \cdots, \tag{14}$$

which obviously can be written in either of the forms

$$\phi_1 \sim \alpha_0(x_1, x_2) + \alpha_1(x_1, x_2)x_3 + \cdots + \alpha_p(x_1, x_2)x_3^p + \cdots, \tag{15}$$

or

$$\phi_1 \sim \alpha_{00}(x_3) + \alpha_{10}(x_3)x_1 + \cdots + \alpha_{mn}(x_3)x_1^m x_2^n + \cdots . \tag{16}$$

We shall now prove

LEMMA 1. *The series* $\alpha_p$ *of* (15) *and the series* $\alpha_{mn}$ *of* (16) *all converge for* $|x_1|$, $|x_2| < r$ *and for* $|x_3| < r$ *respectively.*

Substitute the series (15) into the equation (13) and equate the coefficients of like powers of $x_3$. Evidently $\alpha_0 = x_1$, and $\alpha_1$ satisfies the equation

$$\frac{\partial \alpha_1}{\partial x_1} x_1 m_1 + \frac{\partial \alpha_1}{\partial x_2} x_2 m_2 + \alpha_1(m_3 - m_1 + x_1 P_{30} + x_2 P_{40}) = \Phi_1, \tag{17}$$

where $P_{i0} = P_i(x_1, x_2, 0)$ and $\Phi_1$ is linear in $\alpha_0$ with coefficients which are analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2| \leqq r$, and are zero for $x_1 = x_2 = 0$.

Set

$$\alpha_1 = \beta_0(x_2) + \cdots + \beta_s(x_2)x_1^s + \cdots , \tag{18}$$

where the $\beta_i$ are formal power series in $x_2$. Then $\beta_0$ formally satisfies an equation of the form

$$\frac{d\beta_0}{dx_2} x_2 m_2 + \beta_0(m_3 - m_1 + x_2 P_{400}) = \Phi_{10},$$

where $P_{400}$ and $\Phi_{10}$ are analytic in $x_2$ for $|x_2| \leqq r$. On account of the condition (iii) of the Introduction being satisfied, it readily follows that $\beta_0$ is analytic† in $x_2$ for $|x_2| \leqq r$. In a similar manner it can be shown that all the $\beta_i$ are analytic in $x_2$ for $|x_2| \leqq r$, and if $\alpha_1$ be arranged according to ascending powers of $x_2$ it can be shown in the same manner that the coefficients are analytic in $x_1$ for $|x_1| \leqq r$.

Now define

$$B_s = \beta_0 + \beta_1 x_1 + \cdots + \beta_{s-1} x_1^{s-1}$$

and $\gamma_1 = \alpha_1 - B_s$. Evidently $\gamma_1$ has a factor $x_1^s$, and hence, writing $\gamma_1 = x_1^s \xi_1$, we see that $\xi_1$ will satisfy an equation of the form

$$\frac{\partial \xi_1}{\partial x_1} x_1 m_1 + \frac{\partial \xi_1}{\partial x_2} x_2 m_2 + \xi_1(m_3 - m_1 + s m_1 + x_1 P_{30} + x_2 P_{40}) = \Phi_{11},$$

where $\Phi_{11}$ is analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2| \leqq r$. Let $s$ be large enough so that $m_3 + m_1(s-1)$ is positive and define $m_4$ as equal to this quantity. Let $m$ be a number that is greater than zero and less than the smallest of $m_1$, $m_2$ and $m_4$. Then

$$|\, pm_1 + qm_2 + m_4| > m(p + q + 1)$$

† Cf. II, pp. 177–178.

for any pair of positive integers $p$, $q$. Choose $M>0$ and such that $x_1P_{30}$ and $x_2P_{40}$ are dominated by

$$\frac{M}{\left(1-\frac{x_1}{r}\right)\left(1-\frac{x_2}{r}\right)}-M, \tag{19}$$

and $\Phi_{11}$ is dominated by the first term of (19) only. Then the power series $\xi_1$ is dominated by the power series solution of

$$m\left(x_1\frac{\partial f}{\partial x_1}+x_2\frac{\partial f}{\partial x_2}\right)=f\left[-m+\frac{M}{\left(1-\frac{x_1}{r}\right)\left(1-\frac{x_2}{r}\right)}-M\right]+\frac{M}{\left(1-\frac{x_1}{r}\right)\left(1-\frac{x_2}{r}\right)}. \tag{20}$$

On account of the symmetry in (20) with respect to $x_1$ and $x_2$ the power series solution of (20) converges for $|x_1|$, $|x_2|<R$, where $R$ is the radius of convergence of the power series solution of the differential equation

$$mx_1\frac{dF}{dx_1}=F\left[-m+\frac{M}{\left(1-\frac{x_1}{r}\right)^2}-M\right]+\frac{M}{\left(1-\frac{x_1}{r}\right)^2}.$$

From this it follows that $R=r$. Hence $\alpha_1$ is analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2|<r$.

The proof of the lemma for the other $\alpha_i$ follows readily by induction since every $\alpha_i$ satisfies an equation of the form (17) where $\alpha_1$ is replaced by $\alpha_i$, $m_3$ by $im_3$, and $\Phi_1$ by $\Phi_i$, where $\Phi_i$ is a polynomial in $\alpha_0, \cdots, \alpha_{i-1}$, with coefficients which are analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2|\leqq r$.

The $\alpha_{mn}$ can be proved to be analytic in the same way that the $\beta_i$ of (18) are proved analytic. This completes the proof of Lemma 1.

Similarly, we may prove that when the series corresponding to $\phi_2$ and $\phi_3$ are arranged as power series in $x_3$ the coefficients are analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2|<r$, and when arranged as power series in $x_1$, $x_2$ the coefficients are analytic in $x_3$ for $|x_3|<r$. The only linear term in $\phi_2$ is the one involving $x_2$, and every other term contains as a factor $x_1x_3$ or $x_2x_3$. The coefficient of the linear term is arbitrary and when it is chosen as 1 the series for $\phi_2$ is unique. Let us suppose that this choice has been made. The series for $\phi_3$ has properties similar to those of $\phi_1$ and $\phi_2$ with the additional one that $x_3$ is a factor of every term in the series. We shall choose the coefficient of

the linear term as 1 and this choice uniquely determines all the others.

If the formal series for $\phi_i$, $i=1$, 2, or 3, should converge for $|x_i|<r_0$, $r_0>0$, $i=1, 2, 3$, it can evidently be used to reduce the $i$th differential equation of (12) immediately to the normal form $dz_i^*/dt=m_iz_i^*$.

Even though the series for the $\phi_i$ should diverge, it will be found that they can be "fitted" by functions of class $C^\infty$ in a manner described in

LEMMA 2. *Let the formal series*

$$S(x_1, x_2, x_3) = \sum_{m,n,p=0}^{\infty} \frac{a_{mnp}}{m!n!p!} x_1^m x_2^n x_3^p,$$

*in which all the $a_{mnp}$ are real, have the property that when it is arranged as a power series in $x_3$, i.e.,*

$$S = \alpha_0(x_1, x_2) + \alpha_1(x_1, x_2)x_3 + \cdots + \alpha_p(x_1, x_2)x_3^p/p! + \cdots, \tag{21}$$

*the $\alpha_p$ are analytic in the complex variables $x_1$, $x_2$ for $|x_1|$, $|x_2|\leqq r$, $r>0$, and when it is arranged as a power series in $x_1$, $x_2$, i.e.,*

$$S = \alpha_{00}(x_3) + \alpha_{10}(x_3)x_1 + \cdots + \alpha_{mn}(x_3)x_1^m x_2^n/(m!n!) + \cdots, \tag{22}$$

*the $\alpha_{mn}$ are analytic in the complex variable $x_3$ for $|x_3|\leqq r$. Then there exists a real function $F(x_1, x_2, x_3)$ of the real variables $x_1$, $x_2$, $x_3$ which is continuous together with all its partial derivatives for $|x_1|$, $|x_2|$, $|x_3|\leqq r$, is analytic for $x_1\neq 0$, $x_2\neq 0$, $x_3\neq 0$, satisfies the equations*

$$\frac{\partial^{m+n}F}{\partial x_1^m \partial x_2^n}\bigg|_{x_1=x_2=0} = \alpha_{mn}(x_3), \quad \frac{\partial^p F}{\partial x_3^p}\bigg|_{x_3=0} = \alpha_p(x_1, x_2) \ (m, n, p = 0, 1, \cdots), \tag{23}$$

*and has the property that it and all its partial derivatives are analytic in $x_i$ and $x_j$ for $|x_i|$, $|x_j|\leqq r$, $x_i\neq 0$, $x_j\neq 0$, $x_k=0$, $i\neq j$, $i\neq k$, $j\neq k$, and are analytic in $x_i$ for $|x_i|\leqq r$, $x_j=x_k=0$, $i, j, k=1, 2, 3$.*

Consider the series†

$$F(x_1, x_2, x_3) = \sum_{m,n,p=0}^{\infty} \frac{a_{mnp}}{m!n!p!} x_1^m x_2^n x_3^p (1 - e^{-1/B_{mnp}}), \tag{24}$$

where $i=1$ if $m\geqq n$, $i=2$ if $m<n$, $b_{mnp}=1+|a_{mnp}|$, $B_{mnp}=b_{mnp}x_3^2x_i^2$. At first we shall consider this series only in the six-dimensional complex region defined by

$$|x_j| \leqq r, \quad |\arg x_j| \leqq \theta, \quad 0 < \theta < \pi/8 \qquad (j = 1,2,3). \tag{25}$$

In this region $|\arg x_3^2x_i^2|$ does not exceed $\pi/2$, so that the exponential

† Cf. II, p. 173.

terms $e^{-1/B_{mnp}}$ in the series have exponents $-1/B_{mnp}$ with negative real parts, and so are less than one in absolute value.

Moreover, we find for $|B_{mnp}| \geqq 1$,

$$|a_{mnp}(1 - e^{-1/B_{mnp}})| = |a_{mnp}| \cdot \left| \frac{1}{B_{mnp}} - \frac{1}{2!B_{mnp}^2} + \cdots \right|$$
$$\leqq |a_{mnp}| \cdot (e - 1)/|B_{mnp}| \leqq 2/[|x_3|^2 |x_i|^2].$$

But if, on the contrary, $|B_{mnp}| < 1$, then, by the definition of $B_{mnp}$, $|a_{mnp}| \leqq 1/[|x_3|^2 |x_i|^2]$. Since the quantity in the parentheses in the preceding inequality is obviously less than two in absolute value, we conclude that the inequality holds in this case too and so without restriction on $B_{mnp}$. Thus the given series is dominated by the series

$$\sum_{m,n,p=0}^{\infty} \frac{2}{m!n!p!} |x_1|^{m_1} |x_2|^{n_1} |x_3|^{p-2}, \tag{26}$$

where

$$m_1 = m - 2 \text{ if } m \geqq n \text{ and } m \text{ if } m < n,$$
$$n_1 = n - 2 \text{ if } m < n \text{ and } n \text{ if } m \geqq n. \tag{27}$$

But it follows from the hypotheses concerning the $\alpha_p$ that that part of the series (24) for which $p \leqq 3$ converges uniformly in the region (25); and it follows from the hypotheses concerning the $\alpha_{mn}$ that that part of the series (24) which gives rise to a negative $m_1$ or to a negative $n_1$ converges uniformly in the same region. Hence the series (24) converges uniformly in the region (25), and an examination of this series will readily show that it is analytic except for $x_1 = 0$, $x_2 = 0$, or for $x_3 = 0$, and that

$$F(x_1, x_2, 0) = \alpha_0, \quad F(0, 0, x_3) = \alpha_{00}.$$

In a similar manner, all the series obtained from $F$ by partial differentiation can be shown to have the same properties of analyticity and continuity as has $F$ in the region defined by (25). This is on account of the fact that the $\alpha_p$ of (21) and the $\alpha_{mn}$ of (22) are analytic for $|x_1|, |x_2|, |x_3| \leqq r$ and that for a finite number of differentiations of (24) only a finite number of negative powers of $x_1$ and of $x_2$ are introduced into the dominating series which correspond to these derivative series in the same way that the series (26) does to $F$. We shall consider only $\partial F/\partial x_3$,

$$\frac{\partial F}{\partial x_3} = \sum_{m,n,p-1=0}^{\infty} \frac{a_{mnp}}{m!n!(p-1)!} x_1^m x_2^n x_3^{p-1}(1 - e^{-1/B_{mnp}})$$
$$+ \sum_{m,n,p=0}^{\infty} \frac{a_{mnp}}{m!n!p!} x_1^m x_2^n x_3^p \left( \frac{-2e^{-1/B_{mnp}}}{B_{mnp} x_3} \right).$$

The first of these series is evidently dominated by

$$\sum_{m,n,p-1=0}^{\infty} \frac{2\,|x_1|^{m_1}\,|x_2|^{n_1}\,|x_3|^{p-3}}{m!n!(p-1)!},$$

and the second by

$$\sum_{m,n,p=0}^{\infty} \frac{2\,|x_1|^{m_1}\,|x_2|^{n_1}\,|x_3|^{p-3}}{m!n!p!}.$$

From the remarks made above concerning $\alpha_p$ and $\alpha_{mn}$, etc., it readily follows that $\partial F/\partial x_3$ has the same properties of continuity and analyticity as has $F$ itself. Furthermore,

$$\left.\frac{\partial F}{\partial x_3}\right|_{x_3=0} = \alpha_1.$$

$F(x_1, x_2, x_3)$ is thus a real function in the octant $x_1 \geqq 0$, $x_2 \geqq 0$, $x_3 \geqq 0$, continuous together with all its partial derivatives in this octant, analytic for $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and satisfies the equations (23). Similarly $F$ is seen to have the same properties in each of the other octants and the eight functions thus defined evidently unite to form a function $F(x_1, x_2, x_3)$ defined by the series (24) and having the properties stated in the lemma down to and including those relative to the equations (23). To prove the last statements made in the lemma concerning the function $F$ and its partial derivatives we need merely use dominating series in a manner wholly analogous to that in which they were used above.

## III. Final analytic normalization

The problem in hand is more readily discussed if the functions $Q_1$ and $Q_2$ appearing in the differential equations (12) have the factors $x_2x_3$ and $x_1x_3$, respectively. We shall now show that the assumption that such is the case can be made without the introduction of any further hypotheses than have already been made. To this end let us consider the transformation

$$y_i = s_i(x_1, x_2, x_3) \qquad (i = 1,2,3), \tag{28}$$

where the series $s_i(x_1, x_2, x_3)$ consist of all the terms of the formal series for $\phi_i$, $i=1, 2, 3$, respectively, involving $x_1^m x_2^n$, $m+n=0, 1, \cdots, \lambda+1$, all the terms involving $x_3^p$, $p=0, 1, \cdots, \lambda+1$, and no others. For the present $\lambda$ will be considered as any integer greater than 1 but will be determined in a later section as depending on the $m_i$. The properties of convergence of the coefficients of the formal series for the $\phi_i$ when these series are arranged according to powers of $x_3$ or according to powers of $x_1$ and $x_2$ imply that the series $s_i$ converge for $|x_j| < r$.

From the definition of the $s_i$ it follows that

$$\frac{\partial s_i}{\partial x_1}[x_1(m_1 + x_3P_1) + x_2x_3Q_1] + \frac{\partial s_i}{\partial x_2}[x_2(m_2 + x_3P_2) + x_1x_2Q_2] + \frac{\partial s_i}{\partial x_3}x_3(m_3 + x_1P_3 + x_2P_4) - m_is_i$$

define functions $R_i(x_1, x_2, x_3)$ which have no terms of degree less than $\lambda+2$ in $x_3$ and less than $\lambda+2$ in $x_1$ and $x_2$ together. Since the only linear term in $s_i$ is $x_i$, $i=1, 2, 3$, respectively, when we solve (28) for $x_i=t_i(y_1, y_2, y_3)$ the only linear term in $t_i$ is $y_i$, $i=1, 2, 3$, respectively. Hence the $R_i$ when expressed as functions of $y_j$ have no terms of degree less than $\lambda+2$ in $y_3$ and less than $\lambda+2$ in $y_1$ and $y_2$ together. Thus, when $x_i$ are replaced by an arbitrary solution $x_i(t)$ of (12), the corresponding $y_i$ satisfy the system of differential equations

$$\frac{dy_i}{dt} = m_iy_i + R_i \qquad (i = 1,2,3), \tag{29}$$

where the arguments of the $R_i$ are the $y_j$. But every term in the formal series for $\phi_1$, except the linear term, contains $x_1x_3$ or $x_2x_3$ as a factor, and corresponding statements are true concerning the formal series for the $\phi_2$ and $\phi_3$. This implies that when $x_3=0$, then $y_3=0$, $x_1=y_1$ and $x_2=y_2$, and when $x_1=x_2=0$, then $y_1=y_2=0$ and $x_3=y_3$. Thus the transformation (28) will preserve the form of the differential equations (12) and hence the form of the system (29) is the same as that of the system (12) when the $x$ are replaced by the $y_i$, $i=1, 2, 3$. Furthermore, on account of the properties of the $R_i$ we see that the new $Q_1$ will have $y_2^\lambda y_3^\lambda$ as a factor and the new $Q_2$ will have $y_1^\lambda y_3^\lambda$ as a factor. Hence we are not losing in generality when we assume that the differential equations under discussion have the form

$$\begin{aligned} \frac{dx_1}{dt} &= x_1(m_1 + x_3P_1) + x_2^2x_3^2Q_1, \\ \frac{dx_2}{dt} &= x_2(m_2 + x_3P_2) + x_1^2x_3^2Q_2, \\ \frac{dx_3}{dt} &= x_3(m_3 + x_1P_3 + x_2P_4), \end{aligned} \tag{30}$$

where the $P_i$ and $Q_i$ are analytic in $x_j$ for $|x_j| \leqq r$, $r>0$, $j=1, 2, 3$, $Q_1$ is a function of $x_2$ and $x_3$ only, and $Q_2$ is a function of $x_1$ and $x_3$ only. Furthermore, every $P_i$ has $x_3^{\lambda+1}$ as a factor, and when expanded as a power series in

$x_1$ and $x_2$ contains no term of degree less than $\lambda+1$ in $x_1$ and $x_2$ together; $Q_1$ contains $x_2^\lambda x_3^\lambda$ as a factor, and $Q_2$ contains $x_1^\lambda x_3^\lambda$ as a factor. These facts will be used in the next section. The partial differential equation (13) and the corresponding ones in $\phi_2$ and $\phi_3$ will now be replaced by the system

$$\frac{\partial \phi_i}{\partial x_1}[x_1(m_1 + x_3P_1) + x_2^2 x_3^2 Q_1] + \frac{\partial \phi_i}{\partial x_2}[x_2(m_2 + x_3P_2) + x_1^2 x_3^2 Q_2] + \frac{\partial \phi_i}{\partial x_3}[x_3(m_3 + x_1P_3 + x_2P_4)] = m_i\phi_i \qquad (i = 1,2,3). \tag{31}$$

## IV. Construction of auxiliary invariant surfaces

It happens that if $Q_1=Q_2=0$ we can fairly readily obtain solutions of this system (31) of class $C^\infty$ in a neighborhood of the origin, and thus solve our problem. Hence we shall try to find a transformation of the form

$$x_i = x_i(y_1, y_2, y_3) \qquad (i = 1,2,3), \tag{32}$$

which will reduce this system of partial differential equations to an equivalent system in which the coefficient of $\partial\phi_i/\partial y_j$ has $y_j$ as a factor, $i, j=1, 2, 3$. Evidently we may choose $x_3=y_3$ as the last equation in this transformation. Now if the system of partial differential equations when expressed in terms of the $y_j$ have this additional property, the transformation (32) will take the integral curves of the system of differential equations (30) into the integral curves of a system of differential equations of the form

$$\frac{dy_i}{dt} = y_i(m_i + T_i) \qquad (i = 1,2,3).$$

Hence our search for such a transformation (32) is connected with the search for two invariant surfaces of the system of differential equations (30) other than the known one $x_3=0$. In this connection we shall prove

Theorem 1. *There exists a function $f_1(x_2, x_3)$ which is of class $C^\infty$ in a neighborhood of $x_2=x_3=0$, is analytic in $x_2, x_3$ for $x_2\neq 0$, $x_3\neq 0$, is, together with all its partial derivatives, analytic in either variable when the other variable is zero, and is such that*

$$x_1 = x_2^2 x_3^2 f_1(x_2, x_3)$$

*is an invariant surface. There also exists another function $g_1(x_1, x_3)$ with properties analogous to those of $f_1(x_2, x_3)$ such that*

$$x_2 = x_1^2 x_3^2 g_1(x_1, x_3)$$

*is another invariant surface.*

Since the method to be used for the determination of one of these invariant surfaces can be used for the determination of the other, we shall confine our attention to that surface which will be transformed into the $y_1=0$ plane. To this end consider the transformation

$$y_1 = x_1 - x_2^2 x_3^2 f_1(x_2, x_3), \quad y_2 = x_2, \quad y_3 = x_3, \tag{33}$$

where the function $f_1$ is to be so determined that when the $x_i$ are replaced by an arbitrary solution $x_i(t)$ of (30) the $y_1(t)$ so defined will satisfy a differential equation of the form

$$\frac{dy_1}{dt} = y_1(m_1 + T_1), \tag{34}$$

in which $T_1$ is a function of the $y_i$ of class $C^\infty$ and $T_1(0,0,0)=0$. On differentiating the first equation of (33) with respect to $t$, replacing the derivatives $dx_i/dt$ by their values given in the equations (30) and $x_1$ by its value given in (33), and noting that $y_1=0$ is to be a solution of the resulting differential equation on account of the required form (34), we obtain the result that $f_1(x_2, x_3)$ is a solution of the partial differential equation

$$\frac{\partial f_1}{\partial x_2} x_2(m_2 + x_3 R_1) + \frac{\partial f_1}{\partial x_3} x_3(m_3 + x_2 R_2) = f_1(m_4 - 2x_2 R_2 + x_3 R_3) + Q_1 \tag{35}$$

in which

$$m_4 = m_1 - 2m_2 - 2m_3, \quad R_1 = P_2 + x_2^3 x_3^5 f_1^2 Q_2,$$

$$R_2 = x_2 x_3^2 f_1 P_3 + P_4, \quad R_3 = P_1 - 2R_1,$$

where the $P_i$ and the $Q_2$ are the same functions of $x_2^2 x_3^2 f_1$, $x_2$, $x_3$ as they were of $x_1$, $x_2$, $x_3$ in (30). From the fact that every $P_i$ has $x_3^{\lambda+1}$ as a factor and when expanded as a power series in $x_1$ and $x_2$ contains no term of degree less than $\lambda+1$ in $x_1$ and $x_2$ together, and from the form of the $R_i$ it follows that every $R_i$ has $x_2^{\lambda+1} x_3^{\lambda+1}$ as a factor. If $f_1(x_2, x_3)$ is any solution of the equation (35) which is of class $C^\infty$ in a neighborhood of the origin, the function $y_1$ so defined by the first equation in (33) will satisfy an equation of the form (34) when the $x_i$ are replaced by an arbitrary solution $x_i(t)$ of the differential equations (30). This fact is an immediate consequence of the manner in which the partial differential equation (35) has been derived.

We shall now prove

LEMMA 3. *There exists a solution $f_1(x_2, x_3)$ of the partial differential equation* (35) *which is of class $C^\infty$ in a neighborhood of the origin $x_2=x_3=0$ and is analytic at any interior point of this neighborhood for which $x_2\neq 0$, $x_3\neq 0$. Furthermore, this solution and all its partial derivatives are analytic in either variable in a neighborhood of $x_i=0$, $i=2$ or 3 as the case may be, when the other variable is zero.*

From what has gone before, Theorem 1 will be obvious when Lemma 3 has been proved.

The first thing to be observed is that there exists a unique formal power series solution of the partial differential equation (35) which may be written in any one of the three forms

$$
\begin{aligned}
f_1 &\sim a_{00} + \cdots + \frac{a_{mn}}{m!n!} x_2^m x_3^n + \cdots \\
&= \alpha_0(x_2) + \cdots + \alpha_m(x_2) x_3^m + \cdots \\
&= \beta_0(x_3) + \cdots + \beta_n(x_3) x_2^n + \cdots .
\end{aligned} \tag{36}
$$

Since every $R_i$ and $Q_1$ has $x_2^\lambda x_3^\lambda$ as a factor, this formal series for $f_1$ also has $x_2^\lambda x_3^\lambda$ as a factor. If this series does not diverge for every pair of values $x_2\neq 0$, $x_3\neq 0$, our lemma is obviously proved, so that we shall confine our attention in the following discussion to the contrary case. By making use of the methods employed in the proof of Lemma 1 it can be shown that all the $\alpha_m(x_2)$ are analytic in $x_2$ for $|x_2|<r$ and that all the $\beta_n(x_3)$ are analytic in $x_3$ for $|x_3|<r$. This follows from the fact that wherever $f_1$ appears in $P_i$ or in $Q_2$ it is multiplied by $x_2^2 x_3^2$. Now define a function $g(x_2, x_3)$ which corresponds to the formal series for $f_1$ in the same way that the function $F$ corresponds to the series $S$ in Lemma 2, it being supposed for this application of this lemma that no $x_1$ appears in the series $S$ and that that part of the function $F$ which is independent of $x_2$ is replaced by the corresponding analytic series which is a part of $S$. On making the transformation $f_1=f+g$ we see that $f_1$ satisfies the differential equation (35) if and only if $f$ is a solution of the partial differential equation

$$
\frac{\partial f}{\partial x_2} x_2(m_2 + x_3 S_1) + \frac{\partial f}{\partial x_3} x_3(m_3 + x_2 S_2) = f(m_4 + S_3) + S_4, \tag{37}
$$

in which the $S_i$ are well defined functions of $f, x_2, x_3$, $S_3(0, 0, 0)$ being zero and $S_4$ being independent of $f$. These functions are given explicitly by the following equations (38) and have the properties that they are of class $C^\infty$ in a neighborhood of $f=x_2=x_3=0$, are analytic in $f$ in a neighborhood of $f=0$ for $x_2$ and $x_3$ in a neighborhood of $x_2=x_3=0$, and are analytic in $x_2$, $x_3$ in a

neighborhood of $x_2=x_3=0$ for $x_2\neq 0$, $x_3\neq 0$ and $f$ in a neighborhood of $f=0$. All these properties are immediately implied by the equations

$$S_1(f,x_2,x_3) = R_1(f+g,x_2,x_3),\quad S_2(f,x_2,x_3) = R_2(f+g,x_2,x_3),$$

$$\begin{aligned} S_3(f,x_2,x_3) = &-2x_2R_2 + x_3R_3 - \frac{1}{f}\Big\{\frac{\partial g}{\partial x_2}x_2x_3(R_1-R_{10}) \\ &+\frac{\partial g}{\partial x_3}x_2x_3(R_2-R_{20}) - g[-2x_2(R_2-R_{20}) \\ &+x_3(R_3-R_{30})]\Big\}, \end{aligned} \tag{38}$$

$$\begin{aligned} S_4(f,x_2,x_3) = &Q_1 + g(m_4 - 2x_2R_{20} + x_3R_{30}) - \frac{\partial g}{\partial x_2}x_2(m_2+x_3R_{10}) \\ &-\frac{\partial g}{\partial x_3}x_3(m_3+x_2R_{20}), \end{aligned}$$

where the arguments of the $R_i$ are $f+g$, $x_2$, $x_3$ and $R_{i0}=R_i(0+g,\ x_2,\ x_3)$. Evidently every $S_i$ has $x_2^\lambda x_3^\lambda$ as a factor. This fact will be made use of later on. From the definition of $g(x_2,\ x_3)$ it follows that every partial derivative of $S_4$ is zero when either $x_2$ or $x_3$ is zero. Hence for every pair of positive integers $p$, $q$ a function $S_{pq}$ can be defined which has the same properties of continuity and analyticity that $S_4$ itself has and satisfies the equation

$$S_4 = x_2^p x_3^q S_{pq}. \tag{39}$$

In other words, $S_4$ admits of the factor $x_2^p x_3^q$ for arbitrary $p$ and $q$. This fact can readily be established by taking the factors $x_2$ and $x_3$ out of $S_4$ one at a time, e.g., for $x_2\neq 0$ the function $S_4/x_2$ has all the properties of continuity and of analyticity that $S_4$ itself has. The limits as $x_2\to 0$ of $S_4/x_2$ and all its partial derivatives exist and are zero. Hence $S_4/x_2$ defines $S_{10}$.

Now let us consider the system of ordinary differential equations

$$\begin{aligned} \frac{df}{dt} &= f(m_4+S_3)+S_4, \\ \frac{dx_2}{dt} &= x_2(m_2+x_2S_1), \\ \frac{dx_3}{dt} &= x_3(m_3+x_2S_2). \end{aligned} \tag{40}$$

If we can determine a surface $f=f(x_2,\ x_3)$ in the $f$, $x_2$, $x_3$ space which is made up of integral curves of the system of ordinary differential equations (40), such that $f(x_2,\ x_3)$ is of class $C^\infty$ in a neighborhood of $x_2=x_3=0$, and is analytic at any interior point of this neighborhood for which $x_2\neq 0$, $x_3\neq 0$, the func-

tion $f(x_2, x_3)$ so defined will be a solution of the partial differential equation (37) and will define by means of the equation $f_1=f+g$ a function $f_1(x_2, x_3)$ which will be a solution of the partial differential equation (35). Thus if we can find such an $f(x_2, x_3)$ our lemma will be proved.

Suppose for the sake of definiteness that $m_4$ is greater than zero. We shall limit ourselves to a closed region $R$ about $f=x_2=x_3=0$ in which $|x_i|<1$, $i=2, 3$, $m_4+S_3$ and $m_2+x_3S_1$ are greater than zero, and $m_3+x_2S_2$ is less than zero. We shall suppose that $R$ is further limited so that for $(f, x_2, x_3)$ any point in this region the functions $S_i$ are analytic in $f$, of class $C^\infty$ in $x_2, x_3$, and are analytic in $x_2, x_3$ if $x_2\neq 0$, $x_3\neq 0$. That this is possible follows from the properties of the $S_i$ stated above. The following notation will be adopted for this region $R$:

$$\begin{aligned} &M_1 = \max(m_4 + S_3),\ M_2 = \max(m_2 + x_3S_1),\ M_3 = \max(m_3 + x_2S_2), \\ &N_1 = \min(m_4 + S_3),\ N_2 = \min(m_2 + x_3S_1),\ N_3 = \min(m_3 + x_2S_2). \end{aligned} \tag{41}$$

Now consider the surface defined by the integral curves of the differential equations (40) which pass through the line

$$f = 0,\ x_2 = x_3 = \tau,\ \tau \geqq 0, \tag{42}$$

and suppose that $t=0$ on this line. The equations of this surface may be written as

$$f = f_0(t,\tau),\ x_2 = x_2(t,\tau),\ x_3 = x_3(t,\tau), \tag{43}$$

where the functions appearing on the right hand sides of these equations, on account of the analytic properties of the $S_i$, are analytic in their arguments so long as $\tau\neq 0$. Until further notice, it will be understood that the discussion from now on is relative to that quadrant of $R$ for which $x_2$ and $x_3$ are positive. We wish first of all to show that there is an open region $R_0$ of the $x_2, x_3$ space which is bounded by the axes $x_2=0$, $x_3=0$ and by the arc of a circle, having its center at the origin and its radius different from zero, and is of such a nature that, when $(x_2, x_3)$ is any point in it, the last two equations of (43) can be solved for $t$ and $\tau$ as single-valued functions of $x_2$ and $x_3$. For such a region these functions $t=t(x_2, x_3)$ and $\tau=\tau(x_2, x_3)$ will evidently be analytic, and when they are substituted in the first equation of (43) we shall have $f=f(x_2, x_3)$ where $f(x_2, x_3)$ is an analytic function of its arguments for $(x_2, x_3)$ in $R_0$.

To this end let us find how the $t$ of any point of the surface (43) depends on the corresponding $\tau$. From the second equation of (40) we evidently have

$$\log x_2 - \log \tau = \int_0^t (m_2 + x_3S_1)dt.$$

Since $x_2<1$ on account of the hypotheses concerning the region $R$, for $t\geqq 0$ we may write $-\log\tau\geqq N_2 t$ where $N_2$ is defined in (41). A corresponding inequality exists for $t\leqq 0$, i.e., $-\log\tau\geqq M_3 t$, so that we know that

$$|t| \leqq n|\log\tau|, \tag{44}$$

where $n$ is the greater of $1/N_2$ and $-1/M_3$.

Now for $\tau\neq 0$, and $(f, x_2, x_3)$ in $R$, the partial derivatives $\partial f/\partial\tau$, $\partial x_2/\partial\tau$, and $\partial x_3/\partial\tau$ satisfy the equations of variation

$$\begin{aligned}
\frac{\partial}{\partial t}\cdot\frac{\partial f}{\partial\tau} &= \frac{\partial f}{\partial\tau}(m_4+\omega T_{11}) + \frac{\partial x_2}{\partial\tau}\omega T_{12} + \frac{\partial x_3}{\partial\tau}\omega T_{13},\\
\frac{\partial}{\partial t}\cdot\frac{\partial x_2}{\partial\tau} &= \frac{\partial f}{\partial\tau}\omega T_{21} + \frac{\partial x_2}{\partial\tau}(m_2+\omega T_{22}) + \frac{\partial x_3}{\partial\tau}\omega T_{33},\\
\frac{\partial}{\partial t}\cdot\frac{\partial x_3}{\partial\tau} &= \frac{\partial f}{\partial\tau}\omega T_{31} + \frac{\partial x_2}{\partial\tau}\omega T_{32} + \frac{\partial x_3}{\partial\tau}(m_3+\omega T_{33}),
\end{aligned} \tag{45}$$

where $\omega=x_2^{\lambda-1}x_3^{\lambda-1}$ and the $T_{ij}$, since every $S_i$ has $x_2^{\lambda}x_3^{\lambda}$ as a factor, are defined by the equations

$$\omega T_{11} = S_3 + f\partial S_3/\partial f,\quad \omega T_{12} = f\partial S_3/\partial x_2 + \partial S_4/\partial x_2,\quad \omega T_{13} = f\partial S_3/\partial x_3 + \partial S_4/\partial x_3,$$

$$\omega T_{21} = x_2x_3\partial S_1/\partial f,\quad \omega T_{22} = x_3S_1 + x_2x_3\partial S_1/\partial x_2,\quad \omega T_{23} = \partial(x_2x_3S_1)/\partial x_3,$$

$$\omega T_{31} = x_2x_3\partial S_2/\partial f,\quad \omega T_{32} = \partial(x_3x_2S_2)/\partial x_2,\quad \omega T_{33} = x_2S_2 + x_3x_2\partial S_2/\partial x_3.$$

Now for $t=0$, $\partial f/\partial\tau=0$, $\partial x_2/\partial\tau=\partial x_3/\partial\tau=1$. With this in mind let us set

$$\begin{aligned}
&\xi = \exp\left[\int_0^t (m_2+\omega T_{22})dt\right],\quad \eta = \exp\left[\int_0^t (m_3+\omega T_{33})dt\right],\\
&u = \partial f/\partial\tau,\quad v+\xi = \partial x_2/\partial\tau,\quad w+\eta = \partial x_3/\partial\tau.
\end{aligned} \tag{46}$$

Hence, when $t=0$, $y=v=w=0$. Now make use of the transformation (46) in the equations (45) and obtain

$$\begin{aligned}
\frac{\partial u}{\partial t} &= u(m_4+\omega T_{11}) + (v+\xi)\omega T_{12} + (w+\eta)\omega T_{13},\\
\frac{\partial v}{\partial t} &= u\omega T_{21} + v(m_2+\omega T_{22}) + (w+\eta)\omega T_{23},\\
\frac{\partial w}{\partial t} &= u\omega T_{31} + (v+\xi)\omega T_{32} + w(m_3+\omega T_{33}).
\end{aligned} \tag{47}$$

Hence

$$u = \exp\Big[\int_0^t (m_4 + \omega T_{11})dt\Big]\int_0^t \exp\Big[-\int_0^t (m_4 + \omega T_{11})dt\Big] \cdot [(v+\xi)\omega T_{12} + (w+\eta)\omega T_{13}]dt,$$

$$v = \xi \int_0^t \xi^{-1}[u\omega T_{21} + (w+\eta)\omega T_{23}]dt, \tag{48}$$

$$w = \eta \int_0^t \eta^{-1}[u\omega T_{31} + (v+\xi)\omega T_{32}]dt.$$

We shall limit the discussion for the time being to that part of the first quadrant of the $x_2x_3$ plane which corresponds to $t \geqq 0$ since an argument similar to that about to be given can be made when $t \leqq 0$. Define $U = \max |u|$ for $\tau$ fixed and $t$ variable, and make corresponding definitions for $V$ and $W$. Also define the constants $M_{ij}$ so that $\infty > M_{ij} > n \max |x_2^{\lambda-1} T_{ij}|$ in $R$, where $n$ is that in (44). Since $x_3$ decreases when $t$ increases, we have for this region that $x_3$ is less than its corresponding $\tau$. From the first equation of (48) we obtain

$$\begin{aligned} U &\leqq e^{M_1 t}\big[\{VM_{12} + WM_{13}\}\tau^{\lambda-1} + \{e^{M_3 t}M_{12} + M_{13}\}\tau^{\lambda-1}\big]\,|\log\tau| \\ &\leqq e^{nM_1|\log\tau|}\big[VM_{12} + WM_{13} + e^{nM_2|\log\tau|}M_{12} + M_{13}\big]\tau^{\lambda-1}\,|\log\tau| \\ &\leqq \tau^{-nM_1+\lambda-1}\,|\log\tau|\,\big[VM_{12} + WM_{13} + \tau^{-nM_2}M_{12} + M_{13}\big]. \end{aligned} \tag{49}$$

In a similar manner from the other two equations of (48) we obtain

$$\begin{aligned} V &\leqq \tau^{-nM_2+\lambda-1}\,|\log\tau|\,[\,UM_{21} + WM_{23} + M_{23}], \\ W &\leqq \tau^{nN_3+\lambda-1}\,|\log\tau|\,[UM_{31} + VM_{32} + \tau^{-nM_3}M_{32}]. \end{aligned} \tag{50}$$

For $t \geqq 0$, if $\lambda$ were to satisfy the inequality

$$\lambda > 3 + n(M_1 + M_2 - N_3) \tag{51}$$

we could show that $U$, $V$ and $W$ approach zero as $\tau \to 0$ and hence that the partial derivatives $\partial x_2/\partial\tau$ and $\partial x_3/\partial\tau$ are greater than zero for $\tau$ less than a certain number which is greater than zero.

The inequality which corresponds to (51) for $t \leqq 0$ is the same as (51). Thus it follows from the definition of $n$ and from (51) that it would be sufficient if $\lambda$ were to satisfy the inequality

$$\lambda > 3 + (M_1 + M_2 - N_3)(1/N_2 - 1/M_3). \tag{52}$$

But $M_1$, $M_2$ and $N_3$ depend to some extent on the choice of $\lambda$. Let us choose $\lambda$ so large that

$$\lambda > 3 + L(m_4 + m_2 - m_3)(1/n_2 - 1/m_3)^*$$

* $\lambda$ will also have to satisfy another inequality which arises from the determination of the other invariant surface.

where $L$ is any positive number greater than 1, and then choose $R$ so that (52) holds, which evidently can be done.

But, since $\tau<1$, the max $|\log\tau|\tau$ for the region $R$ is when $\tau=e^{-1}$. Thus for $\tau<e^{-1}$ we may write the inequalities (49) and (50) as

$$
\begin{aligned}
U &\leqq \tau(VM_{12}+WM_{13})+\tau(M_{12}+M_{13}),\\
V &\leqq \tau(UM_{21}+WM_{23})+\tau M_{23},\\
W &\leqq \tau(UM_{31}+VM_{32})+\tau M_{32}.
\end{aligned}
\tag{53}
$$

These inequalities may be written in the form

$$
\begin{aligned}
U-\tau M_{12}V-\tau M_{13}W &\leqq \tau(M_{12}+M_{13}),\\
-\tau M_{21}U+V-\tau M_{23}W &\leqq \tau M_{23},\\
-\tau M_{31}U-\tau M_{32}V+W &\leqq \tau M_{32}.
\end{aligned}
$$

Since the determinant of the coefficients of $U, V, W$ in the left hand members of these inequalities is definitely positive for $\tau$ small enough and since each of the right hand members contains $\tau$ as a factor and the $M_{ij}$ are constants independent of $\tau$, we see that for $\tau$ small enough $U$, $V$ and $W$ are as small as we desire.

This argument, with exception of that part of it relative to the choice of $\lambda$, has been made for $t\geqq 0$. A similar argument can be made for $t\leqq 0$ and hence as is seen from the equations (46), for $\tau$ small enough $\partial x_2/\partial\tau$ and $\partial x_3/\partial\tau$ are greater than zero for $(f, x_2, x_3)$ in $R$ and, as is still assumed, $x_2>0$ and $x_3>0$.

Now examine the jacobian for the last two equations of (43). In $R$ and for $x_2>0$ and $x_3>0$ we have always $\partial x_2/\partial t$ greater than zero and $\partial x_3/\partial t$ less than zero. From these facts and those proved in the preceding paragraphs we see that we can find an open region $R_0$ defined by $x_2>0$, $x_3>0$, $x_2^2+x_3^2<2\tau_0^2$ where $\tau_0$ is a constant greater than zero but so small that the jacobian

$$
\begin{vmatrix}
\partial x_2/\partial\tau & \partial x_3/\partial\tau\\
\partial x_2/\partial t & \partial x_3/\partial t
\end{vmatrix}
$$

is always less than zero. Hence for this region $R_0$ we can always solve the last two equations of (43) for $t$ and $\tau$ as analytic functions of $x_2$ and $x_3$ and thus obtain $f=f(x_2, x_3)$ as has already been described.

We wish now to show that the limits of $f(x_2, x_3)$ and all its derivatives as $x_2\to 0$ or as $x_3\to 0$ are zero. To do this let us note that proving any function $G(x_2, x_3)$ satisfies the relationship

$$
G(x_2, x_3) = O(x_2^p x_3^q) \tag{54}
$$

for an arbitrary pair of positive integers $p$, $q$ is equivalent to proving that it satisfies the relationship

$$G(x_2, x_3) = O(\tau^r) \tag{55}$$

for an arbitrary positive integer $r$. It is understood that $x_2$ and $x_3$ are positive and that $\tau$ is the value of the parameter which corresponds to the point $(x_2, x_3)$ by means of the last two equations of (43).

First of all let us observe that the last two equations of (40) imply that

$$\tau/x_2 \leqq \tau^{-nM_2}, \qquad \tau/x_3 \leqq \tau^{nN_3}. \tag{56}$$

Now consider that part of the quadrant in the $x_2x_3$ plane which corresponds to $t \geqq 0$. For this region $x_3 \leqq \tau$, and the equation (54) implies the existence of a positive constant $M_q$ such that $|G(x_2, x_3)| \leqq x_3^q M_q$, where $q$ is any positive integer. Hence $|G(x_2, x_3)| \leqq \tau^q M_q$. In the same way we can show that for that part of the quadrant which corresponds to $t \leqq 0$ there exists a positive constant $M_p$ corresponding to any arbitrary positive integer $p$ such that $|G(x_2, x_3)| \leqq \tau^p M_p$. If we define $M_{pq}$ as the greater of $M_p$ and $M_q$ we see that for $t$ positive or negative

$$|G(x_2, x_3)| \leqq \tau^r M_{rr} \tag{57}$$

for an arbitrary positive integer $r$, the $M_{rr}$ being a positive constant depending on the $r$. Hence the relationship (54) implies the relationship (55).

Now consider the converse situation. The equation (55) implies the existence of a positive constant $M_{rr}$ and a relationship of the form (57) for an arbitrary positive integer $r$. On using the inequalities (56) we obtain the inequalities

$$\frac{|G|}{x_2^p x_3^q} \leqq \frac{\tau^p}{x_2^p} \cdot \frac{\tau^q}{x_3^q} M_{rr} \tau^{r-p-q} \leqq M_{rr} \tau^{r-p-q-pnM_2+qnN_3}.$$

If we choose $r$ so large that $r-p-q-pnM_2+qnN_3>0$, the quantity after the last inequality sign is finite and hence we have the result that the relationship (55) implies the relationship (54).

We are now in a position to discuss the behavior of $f(x_2, x_3)$ and of its partial derivatives as $x_2 \to 0$ or as $x_3 \to 0$. From the first equation of (40) and from (39) and (42) we have that

$$f = \exp\left[\int_0^t (m_4 + S_3)\,dt\right] \int_0^t \exp\left[-\int_0^t (m_4 + S_3)\,dt\right] O(\tau^r)\,dt,$$

whence

$$|f| \leqq \tau^{-nM_1} |O(\tau^r)|\, n\, |\log \tau| \leqq |O(\tau^{r-nM_3-1})|$$

for $r$ an arbitrary positive integer. Hence $f(x_2, x_3) = O(x_2^p x_3^q)$ for arbitrary positive integers $p$,$q$.

In order to discuss the behavior of $\partial f/\partial x_2$ as $x_2 \to 0$ or as $x_3 \to 0$ we shall make use of the fact that

$$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial t}\cdot\frac{\partial t}{\partial x_2} + \frac{\partial f}{\partial \tau}\cdot\frac{\partial \tau}{\partial x_2}. \tag{58}$$

Since $f(x_2, x_3) = O(x_2^p x_3^q)$ for arbitrary $p$, $q$ as has just been proved, the first equation of (40) shows that $\partial f/\partial t$ has the same property. Now consider the equations (45) once again. From them we have proved that $\partial x_2/\partial \tau$ behaves like $\xi$ of (46) for $\tau$ small enough and hence for $x_2$ or $x_3$ small enough. Hence $\partial x_2/\partial \tau$, as $x_2 \to 0$ or as $x_3 \to 0$, is at most infinite like $\tau^{-nM_2}$. Similarly we have $\partial x_3/\partial \tau = O(\tau^{nN_3})$. Now we are in a position to examine the behavior of $\partial \tau/\partial x_2$ and of $\partial t/\partial x_2$. To do this let us consider the identities

$$\begin{aligned} 1 &= \frac{\partial \tau}{\partial x_2}\cdot\frac{\partial x_2}{\partial \tau} + \frac{\partial \tau}{\partial x_3}\cdot\frac{\partial x_3}{\partial \tau}, \\ 0 &= \frac{\partial \tau}{\partial x_2}\cdot\frac{\partial x_2}{\partial t} + \frac{\partial \tau}{\partial x_3}\cdot\frac{\partial x_3}{\partial t}, \end{aligned} \tag{59}$$

and

$$\begin{aligned} 1 &= \frac{\partial t}{\partial x_2}\cdot\frac{\partial x_2}{\partial t} + \frac{\partial t}{\partial x_3}\cdot\frac{\partial x_3}{\partial t}, \\ 0 &= \frac{\partial t}{\partial x_2}\cdot\frac{\partial x_2}{\partial \tau} + \frac{\partial t}{\partial x_3}\cdot\frac{\partial x_3}{\partial \tau}. \end{aligned} \tag{60}$$

Since the determinant

$$\begin{vmatrix} \partial x_2/\partial \tau & \partial x_3/\partial \tau \\ \partial x_2/\partial t & \partial x_3/\partial t \end{vmatrix}$$

is zero only when $x_2 = x_3 = 0$ and then vanishes like $\tau$, we see that $\partial t/\partial x_2$ and $\partial \tau/\partial x_2$ are at most infinite like a finite power of $\tau^{-1}$ as $x_2 \to 0$ or as $x_3 \to 0$. Let this power be $m$ so that we have that $\partial t/\partial x_2$ and $\partial \tau/\partial x_2$ are $O(\tau^{-m})$ where $m$ is some finite positive number.

Now return to the first differential equation of (45) and note that in addition to the facts used in the discussion there made $T_{12}$ and $T_{13}$ are $O(x_2^p x_3^q)$ for arbitrary positive integers $p$, $q$. On integrating this equation and using the inequalities just established we obtain the result that

$$\left|\frac{\partial f}{\partial \tau}\right| = \left|\exp\left[\int_0^t (m_4 + \omega T_{11})dt\right]\int_0^t \exp\left[-\int_0^t (m_4+\omega T_{11})dt\right] \cdot\left[\frac{\partial x_2}{\partial \tau}\omega T_{12} + \frac{\partial x_3}{\partial \tau}\omega T_{13}\right]dt\right| \leqq \tau^{-nM_1}\,|\,O(\tau^p)\,|\,n\,|\log \tau|\,| = |\,O(\tau^r)\,|$$

for an arbitrary positive integer $r$.

Out of this discussion and the equation (58) follows immediately the fact that $\partial f/\partial x_2 = O(\tau^r) = O(x_2^p x_3^q)$ where $r$, $p$ and $q$ are arbitrary positive integers. It is evident by induction, using the equations (40), (59) and (60), that corresponding statements are true for all the partial derivatives of $f(x_2, x_3)$ with respect to $x_2$ and $x_3$ and hence that all these partial derivatives are zero for $x_2 = 0$ or for $x_3 = 0$.

Although the function $f(x_2, x_3)$ which we have just been discussing has been defined only for the quadrant of $R$ for which $x_2$ and $x_3$ are positive, it is readily seen that it can be defined in the other quadrants in a similar manner and will have the same properties with respect to these quadrants as it has with respect to the first one. When these four functions are joined along the axes they unite to form a function which is of class $C^\infty$ in a neighborhood of the origin, is analytic for $x_2 \neq 0$, $x_3 \neq 0$, and is zero together with all its partial derivatives when $x_2 = 0$ or when $x_3 = 0$. This completes the proof of Lemma 3.

## V. First non-analytic normalization

Thus we have found two invariant surfaces other than $x_3 = 0$. For purposes of notation let the equations of these invariant surfaces be

$$\begin{aligned} x_1 &= x_2^3 x_3^3 f_1(x_2, x_3) = F(x_2, x_3), \\ x_2 &= x_1^3 x_3^3 g_1(x_1, x_3) = G(x_1, x_3). \end{aligned} \tag{61}$$

Reviewing the properties of $F$ and $G$, we know that the function $F(x_2, x_3)$ is of class $C^\infty$ in a neighborhood $|x_2|, |x_3| < r_f, r_f > 0$ of $x_2 = x_3 = 0$ and is analytic at any point $(x_2, x_3)$ of this neighborhood for which $x_2 \neq 0$, $x_3 \neq 0$; the function $G(x_1, x_3)$ is of class $C^\infty$ in a certain neighborhood, $|x_1|, |x_3| < r_g, r_g > 0$, of $x_1 = x_3 = 0$ and is analytic at a point $(x_1, x_3)$ of this neighborhood for which $x_1 \neq 0$, $x_3 \neq 0$. Let $r_0$ be the smaller of $r_f$ and $r_g$.

The aim of this section is to use these invariant surfaces in a transformation which will transform the differential equations (30) into ones of the same form in which the $Q_1 = Q_2 = 0$. Now the invariant surfaces (61) fail to be analytic along the axes and we wish to set up the equations of a transformation which will carry these surfaces into coördinate planes, and will be analytic except possibly for those points which lie on these surfaces. To this end let us examine the transformation

$$\begin{aligned} x_1 &= y_1 + F[x_2 - G(y_1, y_3), y_3], \\ x_2 &= y_2 + G[x_1 - F(y_2, y_3), y_3], \\ x_3 &= y_3, \end{aligned} \tag{62}$$

and confine our attention in the $x_i$ space to a region $R_x$ about the origin for

which $|x_i|<r_0$, and in the $y_i$ space to a region $R_y$ about the origin for which $|y_i|<r_0$. Furthermore, let $R_x$ and $R_y$ be so chosen that the transformation (62) establishes a one-to-one correspondence between the points of $R_x$ and those of $R_y$, which clearly can be done so that the origin $(x_1, x_2, x_3)=(0, 0, 0)$ is interior to $R_x$ and the origin $(y_1, y_2, y_3)=(0, 0, 0)$, which happens to correspond to the origin in the $x_i$ space, is interior to $R_y$. From the properties of $F(x_2, x_3)$ and of $G(x_1, x_3)$ it follows that when $y_1=0$, $x_1=F(x_2, x_3)$, and when $y_2=0$, $x_2=G(x_1, x_3)$. Moreover, this transformation (62) is of class $C^\infty$ and can fail to be analytic only when $y_1=0$, $y_2=0$, or $y_3=0$.

On solving the equations (62) for $x_i$ we obtain

$$x_1 = f_1(y_1, y_2, y_3), \quad x_2 = f_2(y_1, y_2, y_3), \quad x_3 = y_3, \tag{63}$$

where the functions $f_i$ are of class $C^\infty$ in $R_y$ and are analytic at any point of $R_y$ for which $y_1\neq 0$, $y_2\neq 0$, $y_3\neq 0$. Furthermore, in any one of the surfaces $y_1=0$, $y_2=0$, or $y_3=0$, the functions $f_i$ are of class $C^\infty$ in the variables of the surface, analytic at any point which does not lie on either of the axes in the surface, and are analytic along either axis in the variable of the axis in a neighborhood of the origin. This statement is obvious for $y_3=0$. For $y_1=0$, $x_1=F(y_2, y_3)$ and $x_2=y_2$. On account of the properties of the $F(x_2, x_3)$ the statement is seen to be true when $y_1=0$. In a similar manner it can be shown to be true when $y_2=0$.

Since the $f_i$ are of class $C^\infty$ in the $R_y$ we may formally expand these functions as power series in $y_3$ or as power series in $y_1$ and $y_2$ together. Let these expansions be written as

$$\begin{aligned} x_i &\sim \alpha_{i0}(y_1, y_2) + \alpha_{i1}(y_1, y_2)y_3 + \cdots \\ &= \alpha_{i00}(y_3) + \alpha_{i10}(y_3)y_1 + \alpha_{i01}(y_3)y_2 + \cdots, \end{aligned} \tag{64}$$

$i=1, 2$. We shall now prove

LEMMA 4. *The $\alpha_{ij}(y_1, y_2)$ of (64) are analytic in $y_1$, $y_2$ for $|y_1|$, $|y_2|<r_0$, and the $\alpha_{ijk}(y_3)$ are analytic in $y_3$ for $|y_3|<r_0$.*

Since $F(x_2, x_3)$ and $G(x_1, x_3)$ each contains $x_3^2$ as a factor, it is seen that $\alpha_{i0}=y_i$, $i=1, 2$. We shall denote the partial derivative of $F$ with respect to its first argument by $F_1$, and with respect to its second by $F_2$. Furthermore, when the arguments of $F$ are $y_2$, $y_3$ we shall write it merely as $F$, but when the arguments are $x_2-G(y_1, y_3)$, $y_3$ we shall write it as $F^*$. Corresponding notations will be adopted for $G$ and its partial derivatives.

Now differentiate the first two equations of (62) partially with respect to $y_3$. Hence

$$\frac{\partial x_1}{\partial y_3} = F_2^* + F_1^*\left(\frac{\partial x_2}{\partial y_3} - G_2\right), \quad \frac{\partial x_2}{\partial y_3} = G_2^* + G_1^*\left(\frac{\partial x_1}{\partial y_3} - F_2\right),$$

whence

$$\begin{aligned} \frac{\partial x_1}{\partial y_3}(1 - F_1^*G_1^*) &= F_2^* + F_1^*G_2^* - F_1^*G_1^*F_2 - F_1^*G_2, \\ \frac{\partial x_2}{\partial y_3}(1 - F_1^*G_1^*) &= G_2^* + G_1^*F_2^* - G_1^*F_1^*G_2 - G_1^*F_2. \end{aligned} \tag{65}$$

Now for $y_3=0$, $F_1G_1$ and the right hand sides of the equations (65) are zero since $F$ and $G$ both contain $x_3^2$ as a factor. Hence $\partial x_1/\partial y_3$ and $\partial x_2/\partial y_3$ are both zero for $y_3=0$.

Now recall the fact that $F$, as well as all its partial derivatives, is analytic in $x_2$ for $|x_2|<r_0$, $x_3=0$, and that $G$ and its partial derivatives have corresponding properties. Hence any partial derivative of any order of $F^*$ or $G^*$ with respect to the arguments $x_2-G$, $y_3$, and $x_1-F$, $y_3$, respectively, is analytic in $y_2$ and $y_1$, respectively, for $|y_i|<r_0$, $y_3=0$, since $y_1=x_1$ and $y_2=x_2$ for $y_3=0$. The $n$th order partial derivatives of $F_i^*$ and $G_i^*$ with respect to $y_3$ are polynomials in the partial derivatives of $x_1$ and $x_2$ with respect to $y_3$ up to and including those of the $n$th order. The coefficients in these polynomials are polynomials in $F_j^*, G_j^*$, and their partial derivatives with respect to their respective arguments. Hence, from the equations (65) we may determine seriatim all the partial derivatives of $x_1$ and $x_2$ with respect to $y_3$ at $y_3=0$ as functions of $y_1$, $y_2$ which are analytic for $|y_1|$, $|y_2|<r_0$.

Now differentiate the first two equations of (62) partially with respect to $y_1$ and $y_2$. We obtain on simplification

$$\begin{aligned} \frac{\partial x_1}{\partial y_1}[1 - F_1^*G_1^*] &= 1 - F_1^*G_1, \quad \frac{\partial x_2}{\partial y_1} = G_1^*\frac{\partial x_1}{\partial y_1}, \\ \frac{\partial x_2}{\partial y_2}[1 - F_1^*G_1^*] &= 1 - G_1^*F_1, \quad \frac{\partial x_1}{\partial y_2} = F_1^*\frac{\partial x_2}{\partial y_2}. \end{aligned} \tag{66}$$

Now when $y_1=y_2=0$, $x_1=x_2=0$. Hence, by using an argument similar to that used in connection with the partial derivatives with respect to $y_3$, it is seen that from (66) we may determine seriatim all the partial derivatives of $x_1$ and of $x_2$ with respect to $y_1$ and $y_2$ at $y_1=y_2=0$ as functions of $y_3$ which are analytic for $|y_3|<r_0$. This completes the proof of the lemma.

From the equations (65) and (66) we can now prove that all the partial derivatives of $x_1$ and $x_2$, when $y_1$, $y_2$, or $y_3$ is zero, are of class $C^\infty$ in the remaining variables and are analytic in these variables provided neither is zero. We shall prove this only for the $\partial x_1/\partial y_3$ for $y_1=0$ because an induction

proof can be made for all the other derivatives except those of the first order and a proof can be given for them along the lines of that to follow. As we have already seen, if $y_1=0$, then $x_2=y_2$, $x_3=y_3$, and $x_1=F(y_2, y_3)$. Hence the coefficient of $\partial x_1/\partial y_3$ in the first equation of (65) is 1 for $y_1=0$, and it is readily verified that the right hand member of this equation is of class $C^\infty$ in $y_2$ and $y_3$, and is analytic in these variables for $y_2\neq 0$, $y_3\neq 0$.

Now let us perform the transformation of variables defined by the equations (62) or the equations (63). The differential equations (30) will be transformed into differential equations of the same form in $y_i$ since under the transformation (62) the $x_1=x_2=0$ axis and the $x_3=0$ plane are invariant. From the definition of $y_1=0$ it follows that when an integral curve has at any point $y_1=0$, it has $y_1=0$ for all its points in the neighborhood of the origin under consideration. A similar statement is true concerning $y_2$. Hence the right hand members of the first two equations in our new system have $y_1$ and $y_2$, respectively, as factors. Thus the new equations have the form

$$
\begin{aligned}
\frac{dy_1}{dt} &= y_1(m_1 + y_3U_1),\\
\frac{dy_2}{dt} &= y_2(m_2 + y_3U_2),\\
\frac{dy_3}{dt} &= y_3(m_3 + y_1U_3 + y_2U_4),
\end{aligned}
\tag{67}
$$

in which the $U_i$ are functions of the $y_j$ which are of class $C^\infty$ in a neighborhood of the origin and are analytic at any point of this neighborhood for which $y_j\neq 0$, $j=1, 2, 3$. Furthermore, the $U_i$ and all their partial derivatives have the property that in any one of the coördinate planes $y_j=0$, $j=1, 2, 3$, they are of class $C^\infty$ in a neighborhood of the origin, analytic at any point of this neighborhood which does not lie on the coördinate axes of this plane, and analytic along any one of these coördinate axes in the variable of the axis in a neighborhood of the origin. All these statements follow at once from the properties of the transformation (63), because on differentiating the equations of (63) we obtain

$$
\begin{aligned}
\frac{dx_1}{dt} &= \frac{\partial f_1}{\partial y_1}\cdot\frac{dy_1}{dt} + \frac{\partial f_1}{\partial y_2}\cdot\frac{dy_2}{dt} + \frac{\partial f_1}{\partial y_3}\cdot\frac{dy_3}{dt} = x_1(m_1 + x_3P_1) + x_2^2x_3^2Q_1,\\
\frac{dx_2}{dt} &= \frac{\partial f_2}{\partial y_1}\cdot\frac{dy_1}{dt} + \frac{\partial f_2}{\partial y_2}\cdot\frac{dy_2}{dt} + \frac{\partial f_2}{\partial y_3}\cdot\frac{dy_3}{dt} = x_2(m_2 + x_3P_2) + x_1^2x_3^2Q_2,\\
\frac{dx_3}{dt} &= \frac{\partial f_3}{\partial y_1}\cdot\frac{dy_1}{dt} + \frac{\partial f_3}{\partial y_2}\cdot\frac{dy_2}{dt} + \frac{\partial f_3}{\partial y_3}\cdot\frac{dy_3}{dt} = x_3(m_3 + x_1P_3 + x_2P_4).
\end{aligned}
\tag{68}
$$

Now replace the $x_i$ in these equations by the functions of $y_j$ as given in (63). Since all the functions of $y_j$ which now appear in these equations (68) as well as all their partial derivatives have the same properties of continuity and of analyticity that the $f_i$ themselves have, we see at once that our statement concerning the $U_i$ and their partial derivatives is true.

## VI. Second non-analytic normalization

The aim of this section is to reduce the system of differential equations (67) to an equivalent system of the same form for which the formal series for the $\phi_i$ will consist in the linear terms only. To this end we shall first prove

Lemma 5. *Consider the formal series*

$$\begin{aligned}\psi &\sim x_1 + \cdots + \frac{a_{mnp}}{m!n!p!}x_1^m x_2^n x_3^p + \cdots \\ &= x_1 + \alpha_1(x_1, x_2)x_3 + \cdots + \alpha_p(x_1, x_2)x_3^p + \cdots \\ &= x_1 + \cdots + \alpha_{mn}(x_3)x_1^m x_2^n + \cdots\end{aligned} \tag{69}$$

*in which the only linear term is $x_1$, every other term contains $x_1x_3$ or $x_2x_3$ as a factor, $\alpha_p$ are analytic in $x_1$, $x_2$ for $|x_1|$, $|x_2|<r$, $r>0$, and $\alpha_{mn}$ are analytic in $x_3$ for $|x_3|<r$. Into the series (69) substitute the formal series*

$$\begin{aligned}x_i &\sim y_i + \beta_{i1}(y_1, y_2)x_3 + \cdots + \beta_{ip}(y_1, y_2)x_3^p + \cdots \\ &= y_i + \cdots + \beta_{imn}(x_3)y_1^m y_2^n + \cdots, \quad i = 1,2,\end{aligned} \tag{70}$$

*in which the only linear term in the ith series is $y_i$, every other term contains $y_1x_3$ or $y_2x_3$ as a factor, the $\beta_{ip}$ are analytic in $y_1$, $y_2$ for $|y_1|$, $|y_2|<r$ and $\beta_{imn}$ are analytic in $x_3$ for $|x_3|<r$, and arrange the resulting series as*

$$\begin{aligned}\psi &\sim \gamma_0(y_1, y_2) + \cdots + \gamma_p(y_1, y_2)x_3^p + \cdots \\ &= \gamma_{00}(x_3) + \cdots + \gamma_{mn}(x_3)y_1^m y_2^n + \cdots.\end{aligned} \tag{71}$$

*Then $\gamma_i$ are analytic in $y_1$, $y_2$ for $|y_1|$, $|y_2|<r$ and $\gamma_{mn}$ are analytic in $x_3$ for $|x_3|<r$.*

The truth of this lemma is almost obvious. From the properties of the series (69) and (70) it follows at once that $\gamma_0=y_1$ and that $\gamma_{00}=0$. It readily follows that $\gamma_1=\beta_{11}(y_1, y_2)+\alpha_1(y_1, y_2)$ which is obviously analytic for $|y_1|$, $|y_2|<r$. Seriatim, we may prove that all the $\gamma_p$ are analytic for $|y_1|$, $|y_2|<r$, and in a similar manner we may prove that all the $\gamma_{mn}$ are analytic for $|x_3|<r$.

Now consider the effect of the transformation (62) on the formal series

$\phi_i$. From Lemmas 4 and 5 it follows that the series $\phi_i$ when expanded in powers of $y_j$ have the property that when they are arranged as power series in $y_3$ the coefficients are analytic in $y_1$ and $y_2$ for $|y_1|$, $|y_2| < r_0$, and when arranged as power series in $y_1$ and $y_2$ the coefficients are analytic in $y_3$ for $|y_3| < r_0$. But these series in $y_j$ obtained by substituting the series of (64) for the $x_j$ in the series for the $\phi_i$ are the same as the formal series solutions of the partial differential equations

$$\begin{aligned}(72)\qquad &\frac{\partial \phi_i}{\partial y_1} y_1(m_1 + y_3 U_1) + \frac{\partial \phi_i}{\partial y_2} y_2(m_2 + y_3 U_2) \\ &\qquad + \frac{\partial \phi_i}{\partial y_3} y_3(m_3 + y_1 U_3 + y_2 U_4) = m_i \phi_i\end{aligned}$$

since the equations (72) are the transforms of the equations (31) under the transformation (62). From the form of the equations (72) it follows that the formal series for $\phi_i$ has $y_i$ as a factor for $i=1, 2, 3$, respectively. Hence we may write

$$(73)\qquad \phi_i \sim y_i\left(1 + \cdots + \frac{a_{imnp}}{m!n!p!} y_1^m y_2^n y_3^p + \cdots\right) \qquad (i = 1,2,3),$$

in which the series within the brackets have the same properties relative to the convergence of sets of terms as the series for the $\phi_i$ themselves have. Furthermore, the series for $\phi_1$ and $\phi_2$ reduce to $y_1$ and $y_2$, respectively, when $y_3=0$, and the series for $\phi_3$ reduces to $y_3$ when $y_1=y_2=0$. This follows immediately from the form of the partial differential equations (72).

Now "fit" the series within the parentheses appearing in the equations (73) by means of three functions $F_i(y_1, y_2, y_3)$ in the manner described in Lemma 2, leaving the constant terms as they are. Hence we have the transformation

$$(74)\qquad z_i = y_i(1 + F_i(y_1, y_2, y_3)) \qquad (i = 1,2,3),$$

defined in which the $F_i$ have the properties relative to the $\phi_i$ which are peculiar to their mode of definition.

We wish now to determine the type of the system of differential equations which the $z_i$ satisfy when the $y_j$ in the equations (74) are replaced by an arbitrary solution $y_j(t)$ of the differential equations (67). On account of the properties of the power series $\phi_i$ which were mentioned in a preceding paragraph, we see that the transformation (74) has the property that if $y_1=y_2=0$, then $z_1=z_2=0$ and $z_3=y_3$, and if $y_3=0$, then $z_3=0$, $z_1=y_1$ and $z_2=y_2$. Hence the form of the differential equations which the $z_i$ satisfy is the same as the form of the differential equations which the $y_i$ satisfy, since also $y_i=0$ implies that $z_i=0$ for $i=1, 2, 3$.

On remembering that

$$\frac{dz_i}{dt} = \frac{\partial z_i}{\partial y_1} y_1(m_1 + y_3 U_1) + \frac{\partial z_i}{\partial y_2} y_2(m_2 + y_3 U_2) + \frac{\partial z_i}{\partial y_3} y_3(m_3 + y_1 U_3 + y_2 U_4),$$

and on recalling the definition of the $z_j$ we see that

$$\frac{dz_i}{dt} - m_i z_i = R_i \qquad (i = 1,2,3),$$

where the $R_i$ are functions of the $y_j$ which have the same properties as have the $F_i$ in a neighborhood of the origin and are zero together with all their partial derivatives for $y_1 = y_2 = 0$ or for $y_3 = 0$. From these properties of the $R_i$ and those of the transformation (74) it follows that the functions $R_i$ and all their partial derivatives with respect to the $z_j$ are zero when $z_1 = z_2 = 0$ or when $z_3 = 0$.

From this argument follows

LEMMA 6. *Under the transformation* (74) *the system of differential equations* (67) *is equivalent to a system of the type*

$$\begin{aligned} \frac{dz_1}{dt} &= z_1(m_1 + z_3 T_1), \\ \frac{dz_2}{dt} &= z_2(m_2 + z_3 T_2), \\ \frac{dz_3}{dt} &= z_3(m_3 + z_1 T_3 + z_2 T_4), \end{aligned} \tag{75}$$

*where $T_i$ are functions of $z_j$, are of Class $C^\infty$ in a neighborhood of $z_1 = z_2 = z_3 = 0$, are analytic for $z_1 \neq 0$, $z_2 \neq 0$, $z_3 \neq 0$, are analytic together with all their partial derivatives in $z_i$, $z_j$ for $z_k = 0$, $i \neq j$, $i \neq k$, $j \neq k$, and are zero together with all their partial derivatives when $z_1 = z_2 = 0$ or when $z_3 = 0$.*

It may be noted here and will be found of use in a later section that the properties of the $T_i$ imply that each is $O[(z_1^2 + z_2^2)^p z_3^q]$ for any positive integers $p$, $q$.

## VII. FINAL REDUCTION

In this section we shall make the final reduction, i.e., we shall find a transformation which will reduce the equations (75) to the form (8). The finding of such a transformation will be effected when a certain set of three functions is determined. Since all three of these functions can be determined in exactly the same way, we shall limit ourselves to the determination of the first.

In fact we shall prove the following lemma by showing only how to obtain the first of the functions $g_i$:

LEMMA 7. *There exists a transformation of the type*

$$z_i^* = z_i g_i(z_1, z_2, z_3) \qquad (i = 1,2,3),$$

*where the functions* $g_i(z_1, z_2, z_3)$ *are of class* $C^\infty$ *in a neighborhood of the origin and are analytic for* $z_1 \neq 0$, $z_2 \neq 0$, $z_3 \neq 0$, *under which the differential equations* (75) *are equivalent to the differential equations*

$$\frac{dz_i^*}{dt} = m_i z_i^* \qquad (i = 1,2,3).$$

*Furthermore, under this transformation the plane* $z_i = 0$ *corresponds to the plane* $z_i^* = 0$ *for* $i = 1, 2, 3$.

The discussion will be relative to a closed region $S$ of the $z_i$ space which contains the origin as an interior point and in which $z_1^2 + z_2^2 < 1, z_3^2 < 1, m_1 + z_3 T_1$ and $m_2 + z_3 T_2$ are greater than zero and $m_3 + z_1 T_3 + z_2 T_4$ is less than zero. It will be further assumed that $S$ is composed only of points which lie on integral curves of the differential equations of the system (75) which intersect the cone

$$z_1^2 + z_2^2 - z_3^2 = 0. \tag{76}$$

It is readily seen that this hypothesis concerning $S$ is compatible with the others, since, according to the others, the slopes $dz_1/dz_3$ and $dz_2/dz_3$ of any integral curve in $S$ are zero or infinite only for those integral curves lying in the $z_3 = 0$ plane or in the $z_1 = z_2 = 0$ axis. This implies that the integral curves of the differential equations (75) passing through an arbitrary point in a sufficiently small neighborhood of the origin will intersect the cone (76) in $S$. On account of the signs of the left hand members of the equations (75) in $S$, it is seen that every integral curve in $S$ with the exception of those lying in the $z_3 = 0$ plane or in the $z_1 = z_2 = 0$ axis intersects the cone in only one point. It will be found convenient to define $S_0$ as $S$ with the plane $z_3 = 0$ and the axis $z_1 = z_2 = 0$ removed.

If we write the equations of the cone (76) in the parametric form

$$z_1 = \tau_1, \quad z_2 = \tau_2, \quad z_3 = \pm (\tau_1^2 + \tau_2^2)^{1/2}, \tag{77}$$

and choose these equations as the equations giving the initial values of the $z_i$ when $t = 0$, we see that the points in $S_0$ are placed in a one-to-one correspondence with the points in the corresponding $\tau_1$, $\tau_2$, $t$ space by means of the equations of the integral curves of the differential equations (75),

$$z_i = z_i(\tau_1, \tau_2, t) \qquad (i = 1,2,3). \tag{78}$$

where the functions $z_i$ are of class $C^\infty$ in their arguments for $(z_1, z_2, z_3)$ in $S_0$, and are analytic if $\tau_1 \neq 0$, $\tau_2 \neq 0$.

Let us introduce the following notation relative to the region $S$:

$$M_1 = \max(m_1 + z_3T_1),\ M_2 = \max(m_2 + z_3T_2),\ M_3 = \max(m_3 + z_1T_3 + z_2T_4),$$
$$N_1 = \min(m_1 + z_3T_1),\ N_2 = \min(m_2 + z_3T_2),\ N_3 = \min(m_3 + z_1T_3 + z_2T_4). \tag{79}$$

Let $M_0$ be the larger of $M_1$ and $M_2$, and let $N_0$ be the smaller of $N_1$ and $N_2$.

For this region we wish to find how the $t$ of any point of $S_0$ varies with the $\tau_1^2 + \tau_2^2$ of this point. From the first two equations of (75) it follows that

$$\frac{d(z_1^2 + z_2^2)}{z_1^2 + z_2^2} = \frac{2z_1^2(m_1 + z_3T_1) + 2z_2^2(m_2 + z_3T_2)}{z_1^2 + z_2^2}dt.$$

Hence

$$\log(z_1^2 + z_2^2) - \log(\tau_1^2 + \tau_2^2) = \int_0^t \frac{2z_1^2(m_1 + z_3T_1) + 2z_2^2(m_3 + z_3T_2)}{z_1^2 + z_2^2}dt.$$

Since $z_1^2 + z_2^2$ is less than 1 in $S$, $\log(z_1^2 + z_2^2)$ is negative. Consider that portion of $S_0$ which corresponds to $t \geqq 0$. For this portion $z_1^2 + z_2^2$ is greater than $\tau_1^2 + \tau_2^2$. Hence $t \leqq -[\log(\tau_1^2 + \tau_2^2)]\ [1/(2N_0)]$ since $N_0$ is less than the integrand of the integral in the above equation.

By means of a similar argument with respect to the third equation of (75) for $t \leqq 0$ we can show that $t \geqq -[\log(\tau_1^2 + \tau_2^2)]\ [1/(2N_3)]$. If we let $N$ be the smaller of $2N_0$ and $|2N_3|$ we see that

$$|t| \leqq -[\log(\tau_1^2 + \tau_2^2)][1/N], \tag{80}$$

where $(\tau_1, \tau_2, t)$ corresponds to the point $(z_1, z_2, z_3)$ in $S_0$.

Now the first equation of (75) is transformed into one of the same form by any transformation of the type

$$z_1^* = z_1 g_1(z_1, z_2, z_3), \quad z_2^* = z_2, \quad z_3^* = z_3, \tag{81}$$

where $g_1(0, 0, 0)$ is 1. This follows from the fact that this transformation leaves the $z_1 = 0$ plane invariant. Thus we wish to determine the function $g_1(z_1, z_2, z_3)$ so that the corresponding $z_1^*$ will satisfy the equation

$$dz_1^*/dt = m_1 z_1^*, \tag{82}$$

when the $z_i$ in the equations (81) are replaced by an arbitrary solution $z_i(t)$ of (75).

On differentiating the first equation of (81) with respect to $t$ and using the fact that we wish $z_1^*$ to satisfy the equation (82), we obtain

$$\frac{dz_1^*}{dt} = \frac{dz_1}{dt}g_1 + z_1\frac{dg_1}{dt} = z_1 g_1 m_1,$$

whence

$$\frac{d\log g_1}{dt} = -z_3 T_1.$$

If we set $f_1 = \log g_1$, it is evident that, if we can find a function $f_1 = f_1(z_1, z_2, z_3)$ of class $C^\infty$ such that

$$\frac{df_1}{dt} = -z_3 T_1 \tag{83}$$

along every integral curve of (75) and such that $f_1(0, 0, 0)$ is zero, we have solved our problem because in the same way in which we found $f_1$ we can find functions $f_2$ and $f_3$ corresponding to $z_2$ and $z_3$, respectively, in the same way that $f_1$ corresponds to $z_1$.

Now determine the solution

$$f_1 = h_1(\tau_1, \tau_2, t) \tag{84}$$

of the differential equation (83) which corresponds to the solution (78) of (75) and which has the initial value $f_1 = 0$ for $t = 0$. On account of the fact that the right hand members of the differential equations (75) and (83) are of class $C^\infty$ in their arguments and are analytic if $z_i \neq 0$, $i = 1, 2, 3$, the function $h_1$ of (84) is evidently of class $C^\infty$ in its arguments for $\tau_1^2 + \tau_2^2 \neq 0$ and analytic for $\tau_1 \neq 0$, $\tau_2 \neq 0$. Since by means of the equations (78) we can express $\tau_1$, $\tau_2$ and $t$ as single-valued functions of $z_i$ of class $C^\infty$ for $(z_1, z_2, z_3)$ in $S_0$ which will be analytic if $z_i \neq 0$, $i = 1, 2, 3$, we see that we may write the equation (84) as $f_1 = f_1(z_1, z_2, z_3)$ where the function $f_1(z_1, z_2, z_3)$ will be defined only for $(z_1, z_2, z_3)$ in $S_0$, and will be of class $C^\infty$ in its arguments for this region, analytic if $z_i \neq 0$, $i = 1, 2, 3$. We need now only to investigate how this function behaves as $z_3 \to 0$ or as $(z_1, z_2) \to (0, 0)$. We shall prove that it and all its partial derivatives have limits zero as $z_3 \to 0$ or as $(z_1, z_2) \to (0, 0)$ and hence can be so defined that they are of class $C^\infty$ for $(z_1, z_2, z_3)$ in a certain neighborhood of $(0, 0, 0)$.

Let us first observe that, on account of the similarity of the situation here to that in Section IV, it can be readily shown that proving a function $f(z_1, z_2, z_3)$ to be $O[(z_1^2 + z_2^2)^p z_3^{2q}]$ is equivalent to proving it $O[(\tau_1^2 + \tau_2^2)^r]$ where $p$, $q$ and $r$ are arbitrary positive integers.

Now we are in a position to discuss the behavior of $f_1(z_1, z_2, z_3)$. It follows from the equation (83) and from the inequality (80) that

$$|f_1| = \left| \int_0^t O[(z_1^2 + z_2^2)^p z_3^{2q}]dt \right| \leqq |O[(\tau_1^2 + \tau_2^2)^r]| \cdot |\log(\tau_1^2 + \tau_2^2)| \cdot 1/N$$

for an arbitrary positive integer $r$, since $T_1 = O[(z_1^2 + z_2^2)^p z_3^{2q}]$ for arbitrary positive integers $p$, $q$. Hence $f_1(z_1, z_2, z_3) = O[(z_1^2 + z_2^2)^p z_3^{2q}]$ for arbitrary positive integers $p$, $q$.

By the method used in Section IV for the function $f(x_2, x_3)$ it can be shown that every partial derivative of $f_1(z_1, z_2, z_3)$ is also $O[(z_1^2 + z_2^2)^p z_3^{2q}]$ for arbitrary positive integers $p$, $q$, since all the partial derivatives $\partial z_1/\partial \tau_1$, $\partial \tau_1/\partial z_1$, $\partial z_1/\partial t$, etc., are at most infinite like a finite power of $(\tau_1^2 + \tau_2^2)^{-1}$. Hence the function $f_1(z_1, z_2, z_3)$ is of class $C^\infty$ in a neighborhood of the origin and therefore, on account of its definition, the function $g_1(z_1, z_2, z_3)$ of (81) has the same property. Since the analytic properties of the $f_1$ also imply that the $g_1$ has corresponding analytic properties, and since the other functions $g_2$ and $g_3$ can be determined in the same way in which the $g_1$ has been determined and will have corresponding properties of continuity and of analyticity, we see that the proof of Lemma 7 is complete.

We may summarize the discussion of this and the preceding sections in the concluding

THEOREM 2. *There exists a transformation of the type*

$$x_i = x_i(z_1^*, z_2^*, z_3^*) \qquad (i = 1,2,3), \tag{85}$$

*where the functions $x_i(z_1^*, z_2^*, z_3^*)$ are of class $C^\infty$ in some neighborhood $(z_1^*, z_2^*, z_3^*) = (0, 0, 0)$ and are analytic for $z_1^* \neq 0$, $z_2^* \neq 0$, $z_3^* \neq 0$, under which the differential equations* (7) *are equivalent to the differential equations*

$$\frac{dz_i^*}{dt} = m_i z_i^* \qquad (i = 1,2,3),$$

*in a certain neighborhood of the origin, which point remains invariant under the transformation. Furthermore, when the equations* (85) *are solved for the $z_i^*$ as functions of the $x_j$ and are expanded as formal power series in the $x_j$, if the equations* (7) *have the form* (12) *the formal power series for $z_1^*$ will formally satisfy the partial differential equation* (13), *and the formal power series for $z_2^*$ and $z_3^*$ will formally satisfy the analogues of* (13) *in $\phi_2$ and $\phi_3$.*

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# CUT POINTS OF CONNECTED SETS AND OF CONTINUA*

BY

G. T. WHYBURN

1. **Introduction.** In this paper it will be shown, among other things, that the set $G$ of all the cut points of any continuum $M$ in a locally compact, metric, and separable space is a $G_{\delta\sigma}$, i.e., the sum of a countable number of $G_\delta$ sets.† From this result, with the aid of a well known theorem of W. H. Young's, it follows that if the set $G$ is uncountable it must contain a perfect set.‡ Thus the set of all cut points of any continuum is either vacuous, finite, countable, or of the power of the continuum.

The customary notation and terminology of point set theory will be employed. For example, $\overline{X}=X+X'$, where $X'$ is the set of all limit points of the set $X$; $K \subset H$ means that $K$ is a subset of $H$ or that $H$ contains $K$; $K \cdot H$ denotes the set of points common to $K$ and $H$; $\rho(X, Y)$ denotes the distance from $X$ to $Y$ when $X$ and $Y$ are points, and denotes the minimum distance between $X$ and $Y$ when $X$ and $Y$ or either contains more than one point, i.e. the greatest lower bound of the aggregate of numbers $[\rho(x,y)]$, where $x$ and $y$ are points of $X$ and $Y$ respectively; and $\delta(M)$ denotes the diameter of the set $M$, i.e., the least upper bound of the aggregate $[\rho(x,y)]$, where $x$ and $y$ are two points of $M$. In addition, if $X$ is a cutting of a connected set $M$, unless otherwise stated, the equation $M-X=M_a(X)+M_b(X)$ is to be interpreted as meaning that $M-X$ is the sum of the two mutually separated sets $M_a(X)$ and $M_b(X)$ containing the sets $A$ and $B$ respectively. When this equation is true, $X$ is said to separate $A$ and $B$ in $M$.

All point sets considered are assumed to lie in a locally compact, metric, and separable space.

2. **Cut points and regular points.** We prove the following theorem:

---

* Presented to the Society, December 27, 1928. Received by the editor of the Bulletin in December, 1928, accepted for publication in the Bulletin, and subsequently transferred to the Transactions.

† It has been shown by C. Zarankiewicz that the set of all cut points of any *continuous curve* is an $F_\sigma$ but that this is not true of continua in general; see *Sur les points de division dans les ensembles connexes*, Fundamenta Mathematicae, vol. 9 (1927), pp. 124–171; see pp. 163, 164.

‡ According to results due to F. Bernstein (Leipziger Berichte, vol. 60 (1908), p. 325), there exist uncountable sets, even uncountable linear sets, which contain no perfect set. The theorems in this paper show, however, that such a set cannot be the set of cut points of any continuum.

THEOREM 1. *If $A$ and $B$ are any two mutually exclusive closed subsets of a connected set $M$ and $K$ is the set of all those points of $M$ which separate $A$ and $B$ in $M$, then $K+A+B$ contains every point $P$ of $\overline{K}$ which is a regular point* of $M+P$.*

Let $P$ be any limit point of $K$ which is not in $A+B$ and which is a regular point of $M+P$. As shown in my paper *Concerning connected and regular point sets*† either there exists a sequence $X_1, X_2, X_3, \cdots$ of points of $K$ having $P$ as its sequential limit point and such that, for every positive integer $i$,

$$M - X_i = M_a(X_i) + M_b(X_i), \quad \sum_{n=1}^{i-1} X_n \subset M_a(X_i), \quad \text{and} \quad \sum_{n=i+1}^{\infty} X_n \subset M_b(X_i),$$

or such a sequence exists such that, for each $i$,

$$\sum_{n=1}^{i-1} X_n \subset M_b(X_i) \quad \text{and} \quad \sum_{n=i+1}^{\infty} X_n \subset M_a(X_i).$$

The two cases are alike so we shall consider only the former. Let $E = \sum M_a(X_i)$, and let $F = M - E$. Then $F = \Pi\, M_b(X_i)$, i.e. $F$ is identically the set of points common to all the sets $[M_b(X_i)]$. Thus no point of $E$ is a limit point of $F$, because each point of $E$ belongs to a set $M_a(X_k)$, and $F \subset M_b(X_k)$ for every $k$. I shall now show that no point of $F$ except possibly $P$ (in case $P$ belongs to $M$) is a limit point of $E$. Suppose, on the contrary, that such a limit point $Q$ of $E$ does exist. Let $\epsilon$ be a positive number which is $< \rho(P, Q+A+B)$. Since $P$ is a regular point of $M+P$, there exists a finite subset $U$ of $M$ which $\epsilon$-separates $P$ in $M+P$, i.e., $(M+P)-U = M_p+M_0$, where $M_p$ and $M_0$ are separated, $M_p \supset P$, and $\delta(M_p) < \epsilon$. Now‡ $M_0+U$ is the sum of a finite number of mutually separated connected point

* The point $P$ of a connected point set $M$ is called a Menger regular point of $M$, or simply a regular point of $M$, if for each $\epsilon > 0$, $P$ can be $\epsilon$-separated in $M$ by some finite subset $U$ of $M$, i.e., a finite subset $U$ of $M$ exists such that $M - U = M_p(U) + M(U)$, where $M_p(U)$ and $M(U)$ are mutually separated, $M_p(U) \supset P$, and $\delta[M_p(U)] < \epsilon$. If the $\epsilon$-separating set $U$ can, for every $\epsilon > 0$, be chosen of power $\leqq n$ but cannot, for every $\epsilon$, be chosen of power $< n$, then $P$ is a point of order $n$ of $M$. See K. Menger, *Grundzüge einer Theorie der Kurven*, Mathematische Annalen, vol. 95 (1925), pp. 277–306, and P. Urysohn, *Sur la ramification des lignes Cantoriennes*, Comptes Rendus, vol. 175 (1922), p. 481.

† Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 685–689, see proof of Theorem 1. The proof here given is somewhat similar to the one just referred to. R. L. Wilder has recently published some interesting extensions of Theorem 1 in my paper here cited; see R. L. Wilder, *On connected and regular point sets*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 649–655.

‡ See Knaster and Kuratowski, *Remark on a theorem of R. L. Moore*, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 647–649; see also an abstract of mine in the Bulletin of the American Mathematical Society, vol. 33 (1927), p. 388.

sets $E_1, E_2, E_3, \cdots, E_n$. One of these, say $E_k$, contains $Q$, and $Q$ is not a limit point of $M-E_k$. But there exists an integer $m$ such that for every integer $j>m$, $X_j \subset M_p$. And since $Q \subset M_b(X_j)$, $E_k \supset Q$, and $E_k \cdot X_j = 0$, then $E_k \subset M_b(X_j)$ for every $j>m$. Hence $Q$ is not a limit point of $\sum_{j=m+1}^{\infty} M_a(X_j)$, because this set of points is a subset of $M-E_k$. But clearly $Q$ is not a limit point of $\sum_{j=1}^{m} M_a(X_j)$, because $Q \subset F \subset M_b(X_j)$ for every $j$. Therefore $Q$ is not a limit point of $E$, contrary to supposition. Thus no point of $F$ except possibly $P$ can be a limit point of $E$.

Now $P$ must belong to $F$, for if not, then $E$ and $F$ are mutually separated; and since $E+F=M$, this contradicts the fact that $M$ is connected. Hence $P \subset F$; and since $E$ and $F-P$ are mutually separated and contain $A$ and $B$ respectively, and $E+(F-P)=M-P$, therefore $P$ separates $A$ and $B$ in $M$ and hence belongs to $K$.

COROLLARY 1a. *If $M$ is a continuum, then every point of $\overline{K}$ which is not in $K+A+B$ is a non-regular point of $M$.*

Theorem 1 does not hold true, even in case $M$ is a continuum, when we substitute the words "point of connectivity im kleinen" for the words "regular point." For let $I$ be the interval (0,2) of the $X$-axis; let $A$ and $B$ be the end points of $I$; let $L$ be the straight line interval from (1,0) to (1,1); let $P$ be the point (1,0); for each positive or negative integer $n$, let $L_n$ be the straight line interval from $(1+1/(2n),0)$ to $(1+1/(2n),1)$; and let $M$ be the continuum $I+L+\sum L_n$. Then the set $K$ of points of $M$ which separate $A$ and $B$ in $M$ is identical with the set $I-(A+B+P)$; and $P$ is a limit point of $K$ and is a point of connectivity im kleinen of $M$; but $P$ does not belong to $K$.

THEOREM 2. *Let $N$ be any closed subset of a connected point set $M$, and let $K$ be the set of all points of $M$ which separate* $N$ in $M$. Then $K+N$ contains every point $P$ of $\overline{K}$ which is a regular point of $M+P$.*

Let $P$ be any point of $\overline{K}$ which is not in $N$ and which is a regular point of $M+P$, let $\epsilon$ be a positive number which is $<\rho(P,N)$, and let $U$ be a finite subset of $M$ which $\epsilon$-separates $P$ in $M+P$. Then $M-U=M_p+M_n$, where $M_p$ and $M_n$ are separated and contain $P$ and $N$ respectively. Now $M_p$ contains an infinite sequence $X_1, X_2, X_3, \cdots$ of points of $K$ having $P$ for its sequential limit point. For each $i$, $M-X_i=M_1(X_i)+M_2(X_i)$, where $M_1(X_i)$ and $M_2(X_i)$ are separated and $N \cdot M_1(X_i) \neq 0 \neq N \cdot M_2(X_i)$. Now since† $M_n+U$ is the sum of a finite number of connected point sets, it follows that there exist two point sets $K_a$ and $K_b$ and an infinite subsequence

* The point $P$ of a connected set $M$ is said to separate a given subset $N$ of $M$ in $M$ if $M-P$ is separated between some two points of $N$, i.e., $M-P$ is the sum of two mutually separated sets $M_1$ and $M_2$, where $N \cdot M_1 \neq 0 \neq N \cdot M_2$; see R. L. Wilder, loc. cit.

† See the reference to Knaster and Kuratowski above.

$X_{n_1}, X_{n_2}, \cdots$ of the sequence $[X_n]$ such that $K_a+K_b=M_n+U$ and such that, for every $i$, $K_a \subset M_1(X_{n_i})$ and $K_b \subset M_2(X_{n_i})$. Hence if $A=N \cdot K_a$ and $B=N \cdot K_b$, then $A$ and $B$ are mutually exclusive, closed, and non-vacuous subsets of $M$, and every point of $\sum X_{n_i}$ separates $A$ and $B$ in $M$. Then since $P$ is a limit point of $\sum X_{n_i}$ and is a regular point of $M+P$, it follows by Theorem 1 that $P$ belongs to $M$ and separates $A$ and $B$ in $M$. Therefore $P$ separates $N$ in $M$ and hence belongs to $K$. This completes the proof.

THEOREM 3. *Let $A$ and $B$ be any two mutually exclusive closed subsets of a continuum $M$, and let $K$ be the set of all those points of $M$ which separate $A$ and $B$ in $M$. Then $K+A+B$ contains every point $P$ of $\overline{K}$ having the property that every subcontinuum of $M$ containing $P$ contains at least two points at which $M$ is connected im kleinen.*

Let $P$ be a point of $\overline{K}$ not in $A+B$ which has the property mentioned in this theorem. Let the points and point sets $X_1, X_2, X_3, \cdots, E$, and $F$ be selected and defined exactly as in the proof of Theorem 1. Then, just as in that proof, no point of $E$ is a limit point of $F$. Likewise no point of $F-P$ can be limit point of $E$. For suppose some point $Q$ of $F-P$ is a limit point of $E$. Let $R$ be a compact neighborhood of $P$ such that $\overline{R}$ does not contain $Q$. Then since for no $i$ can $Q$ be a limit point of $M_a(X_i)$, it readily follows that if, for each $i$, $I(X_i,X_{i+1})$ denotes the set of points $M_b(X_i) \cdot M_a(X_{i+1})+X_i+X_{i+1}$, then both $Q$ and $P$ belong to the sequential limiting set $L$ of the sequence of continua* $[I(X_i,X_{i+1})]$. Now with the aid of a theorem of Janiszewski's† and a theorem of Lubben's,‡ it follows that $L \cdot \overline{R}$ contains a continuum $H$ containing $P$ and at least one point of the boundary of $R$. Since for each $i$, $L \cdot M_a(X_i)=0$, clearly $H \subset F$. Now by hypothesis, $H$ contains a point $C$, distinct from $P$, at which $M$ is connected im kleinen. Since $C$ belongs to $L$, it follows that there exists an integer $k$ and a point $D$ of $I(X_k,X_{k+1})$ which can be joined to $C$ by a subcontinuum $I$ of $M$ which does not contain $P$. There exists an integer $j>k$ such that $X_j$ does not belong to $I$. But then $C$ belongs to $M_b(X_j)$ (for $C \subset F \subset M_b(X_i)$, for every $i$), $D$ belongs to $M_a(X_j)$, and $I$ is a connected subset of $M_a(X_j)+M_b(X_j)$ containing both $C$ and $D$. Clearly this is impossible, because $M_a(X_j)$ and $M_b(X_j)$

* See my paper *Concerning the cut points of continua*, these Transactions, vol. 30 (1928), pp. 597–609, Theorem 1.

† Z. Janiszewski, *Sur les continus irréductibles entre deux points*, Journal de l'Ecole Polytechnique, (2), vol. 16 (1912), p. 109.

‡ R. G. Lubben, *Concerning limiting sets in abstract spaces*, these Transactions, vol. 30 (1928), pp. 668–685; see Theorems 11 and 12.

are mutually separated. Thus the supposition that $F-P$ contains a limit point of $E$ leads to a contradiction. Hence $E$ and $F-P$ are mutually separated; and since $E \supset A$, $F-P \supset B$, and $E+(F-P)=M-P$, it follows that $P$ separates $A$ and $B$ in $M$ and hence belongs to $K$.

3. **Cardinal number and Borel classification of the set of cut points.** We shall prove the following theorem:

THEOREM 4. *Let $A$ and $B$ be any two points, or, indeed, any two closed mutually exclusive subsets, of a continuum $M$, and let $K$ be the set of all points of $M$ which separate $A$ and $B$ in $M$. Then $K$ is the sum of a $G_\delta$ set* and a countable set.*

Let $E$ be the set of all those points of $M$ which can, for each $\epsilon>0$, be $\epsilon$-separated in $M$ by some two points of $K$, i.e., the set of all points of $M$ of order $\leqq 2$ relative to $K$. It follows readily with the aid of a theorem of Menger's† that if $E$ exists it is a $G_\delta$ set. And since $A+B$ is closed, $H=E-E\cdot(A+B)$ is a $G_\delta$ set. Now since every point of $E$ is a limit point of $K$ and is a regular point of $M$, then by Theorem 1, $K+A+B \supset E$. Hence $H \subset K$. And by a theorem of the author's,‡ $K-H$ is countable. This completes the proof.

COROLLARY§ 4a. *If $K$ is uncountable, it contains a perfect set.*

THEOREM 5. *The set $G$ of all the cut points of any continuum $M$ is a $G_{\delta\sigma}$, i.e., the sum of a countable number of $G_\delta$ sets.*

Since our space is metric and separable, $M$ itself is‖ separable and hence contains a countable set $D$ such that $M=\bar{D}$. Let $H$ be the collection of all possible pairs of points of $D$, and for each pair $(A,B)$ of points in $H$, let $K_{ab}$ be the set of all points of $M$ which separate $A$ and $B$ in $M$. By Theorem 4, $K_{ab}=$ a $G_\delta$ set + a countable set, for every pair $(A,B)$ in $H$. Then since $H$ is countable, it follows that $\sum_H K_{ab}$ is a $G_{\delta\sigma}$. Now if $X$ is any point of $G$, $M-X=M_1+M_2$, where $M_1$ and $M_2$ are mutually separated. And if $P_1$ and $P_2$ are points of $M_1$ and $M_2$, respectively, then since $P_1\cdot M_2'=0=P_2\cdot M_1'$,

---

* A $G_\delta$ set is a point set which is the common part of some family of open sets.

† See K. Menger, loc. cit., Theorem 3.

‡ See *Concerning the cut points of continua*, loc. cit., Theorem 7; see also my paper *Concerning collections of cuttings of connected point sets*, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 87–104, Theorem 12.

§ This corollary follows from Theorem 4 and Young's theorem that every $G_\delta$ set containing a subset which is dense in itself contains a perfect set; see W. H. Young, Leipziger Berichte, vol. 55 (1903), p. 287.

‖ See W. Gross, *Zur Theorie der Mengen, in denen ein Distanzbegriff definiert ist*, Wiener Sitzungsberichte, vol. 123 (1924), pp. 801–819.

and $P_1+P_2 \subset \overline{D}$, then $D$ is a subset of neither $M_1+X$ nor $M_2+X$. Hence there exist points $A$ and $B$ of $D$ in $M_1$ and $M_2$ respectively. Therefore $X$ separates $A$ and $B$ in $M$ and hence belongs to $K_{ab}$ and to $\sum K_{ab}$. Thus $G \equiv \sum K_{ab}$, and hence $G$ is a $G_{\delta\sigma}$.

COROLLARY* 5a. *If $G$ is uncountable, it contains a perfect set.*

We note here the following fact which, in a certain sense, is a converse proposition to Theorem 4:

*If $K$ is any set which lies within the linear interval $I(A, B)$ and is the sum of a $G_\delta$ set and a countable set, then there exists a continuum $M$ such that $K$ is identically the set of all those points of $M$ separating $A$ and $B$ in $M$.*

We note also the fact that the set of all the cut points of a continuum is not necessarily the sum of a $G_\delta$ set and a countable set. Indeed, *the set of all cut points of the universal acyclic continuous curve of Wazewski is not the sum of a $G_\delta$ set and a countable set.*

4. **Concluding remarks.** Theorem 5 and Corollary 5a are propositions concerning the set $G$ of all the cut points of a continuum $M$. We may obtain the same conclusions for the set $H$ of all the *local separating points*† of a continuum $M$. We state the following propositions concerning the local separating points of continua.

(i) Every continuum $M$ contains a countable collection $[M_i]$ $(i=1,2,3, \cdots)$ of compact subcontinua such that if $H$ is the set of all local separating points of $M$ and, for each $i$, $G_i$ is the set of all cut points of $M_i$, then (1) $\sum_1^\infty G_i \subset H$, (2) $H-\sum G_i$ is countable, and (3) for each $i$, $M_i \cdot (M-M_i)$ contains exactly two points.

(ii) The set $H$ of all the local separating points of any continuum $M$ is a $G_{\delta\sigma}$; and if $H$ is uncountable, it contains a perfect set.

(iii) Every continuous curve $M$ contains a countable collection $[M_i]$, $(i=1,2,3, \cdots)$ of compact continuous curves such that if $H$ is the set of all

---

* This corollary can be deduced from Theorem 5 using Young's theorem (loc. cit.), or it follows also from Corollary 4a and the author's theorem (see *Concerning the cut points of continua*, loc. cit., Theorem 3) that if $G$ is any uncountable set of cut points of $M$, then some two points $A$ and $B$ of $G$ are separated in $M$ by uncountably many points of $G$.

† The point $P$ of a continuum $M$ is a local separating point of $M$ if there exists a compact neighborhood $R$ of $P$ such that $M \cdot \overline{R-P}$ is separated between some two points of the component of $M \cdot \overline{R}$ which contains $P$; see my paper *Local separating points of continua*, in Monatshefte für Mathematik und Physik, vol. 35, No. 2.

local separating points of $M$ and, for each $i$, $G_i$ is the set of all cut points of $M_i$, then $\sum_1^\infty G_i \equiv H$.

Result (i) is readily deduced from the theorems in my paper *Local separating points of continua* (loc. cit.)—in particular from the result that all save possibly a countable number of the local separating points of any continuum $M$ are points of order two of $M$. Result (ii) follows immediately from (i) and Theorem 5 and Corollary 5a above. Result (iii) was established incidentally by the author in proving the theorem* that the set $H$ of all the local separating points of any continuous curve is an $F_\sigma$.

It has been shown by the author† that all save possibly a countable number of the cut points of any *continuum* $M$ are points of order two of $M$. That this theorem does not hold true for *connected sets* $M$, even in case $M$ is irreducibly connected between some two of its points, is shown in the example given by Vietoris‡ of a totally discontinuous function $y=\omega(x)$ which has a connected graph. That this theorem does not hold true even for *connected and connected im kleinen sets* $M$ is shown by the following example. Let $K$ be a non-dense perfect set on the interval $I(0,1)$ of the $X$-axis not containing the end points $A$ and $B$ of $I$. Let $S_1, S_2, S_3, \cdots$ be the complementary segments of $K$ in $I$. For each $i$, let $E_i$ be an ellipse with minor axis $\overline{S}_i$ and with major axis two units in length. Let $M = I + \sum_1^\infty E_i$. Then $M$ is connected and connected im kleinen, and every point of $K$ separates $A$ and $B$ in $M$; but clearly no point of $K$ is a regular point of $M$.

Recently I have made a study of the structure of connected and connected im kleinen point sets and, among other results, have found the following theorem:

*In a connected and connected im kleinen point set $M$, every connected set of cut points is arcwise connected.*§

---

* See my paper *Concerning points of continuous curves defined by certain im kleinen properties* Mathematische Annalen, vol. 102 (1929), pp. 133–336.

† See *Concerning the cut points of continua*, loc. cit., Theorem 7.

‡ L. Vietoris, *Stetige Mengen*, Monatshefte für Mathematik und Physik, vol. 31 (1921), p. 202; see also Knaster and Kuratowski, *Sur quelques propriétés topologiques des fonctions dérivées*, Rendiconti del Circolo Matematico di Palermo, vol. 49 (1925).

§ For special cases of this theorem, see an abstract by C. M. Cleveland in the Bulletin of the American Mathematical Society, vol. 32 (1926), p. 420, where the theorem is stated for *plane continuous curves* $M$, and my paper *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), pp. 167–194, Theorem 8, where the theorem, in more general form, is proved for continuous curves $M$ in $n$-dimensional space.

It is of interest in this connection to note the theorem, easily deducible from the work of F. Frankl*, that *Every connected point set $M$ each point of which is of order $\leqq 2$ of $M$ is arcwise connected.*

* See *Ueber die zusammenhängenden Mengen von höchstens zweiter Ordnung*, Fundamenta Mathematicae, vol. 11 (1927), pp. 96–104.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

# ON A CERTAIN RING OF FUNCTIONS OF TWO VARIABLES*

BY

J. F. RITT

The product of two functions of the type

$$a_1(x)\alpha_1(y) + \cdots + a_n(x)\alpha_n(y) \tag{1}$$

is a function of similar type. There arises thus the problem of determining all representations of any function (1) as a product of functions (1).† The present paper throws some light upon this problem, without coming near to a complete solution of it.

We shall limit ourselves to functions with analytic $a$'s and $\alpha$'s.

We write each function (1) in such a way that the number of terms, $n$, in it is a minimum. This is accomplished by getting rid of any linear dependence which may exist among the $a$'s and among the $\alpha$'s. When the expression (1) for a function has a minimum number of terms, we say that the function is in *reduced* form. The expression in reduced form is unique to within linear transformations upon the $a$'s and $\alpha$'s.

Our first result states that *if* $A = BC$, *where*

$$A = \sum_{i=1}^{n} a_i(x)\alpha_i(y), \quad B = \sum_{i=1}^{r} b_i(x)\beta_i(y), \quad C = \sum_{i=1}^{s} c_i(x)\gamma_i(y),$$

*with* $A$, $B$, $C$ *in reduced form, then every* $b_ic_j$ *is an algebraic combination of the* $a$*'s*; that is, every $b_ic_j$ satisfies an algebraic equation whose coefficients are polynomials in the $a$'s with constant coefficients. Similarly, every $\beta_i\gamma_j$ is algebraic in the $\alpha$'s.

It follows immediately that the ratio of any two $b$'s, or of any two $c$'s, is algebraic in the $a$'s. If, for instance, $b_1 = 1$, a situation which can be brought about by multiplying $B$, and dividing $C$, by a function of $x$, then every $b$ and every $c$ will be algebraic in the $a$'s. *In short, in factoring a function* (1) *into similar functions, no new transcendencies are introduced.*

It is a consequence of the above result that if a function has only one term, it cannot be the product of two functions, either of which, in reduced form, has more than one term.

* Presented to the Society, October 26, 1929; received by the editors November 8, 1929.

† I have learned, in conversation with Professor Blichfeldt, that he also had thought of this question, and, in fact, quite a number of years ago.

The algebraic dependence theorem reduces the general factorization problem to a problem in algebraic functions of several variables. It becomes a question of determining all representations of a function

$$\sum_{i=1}^{n} a_i(u_1, \cdots, u_p)\alpha_i(v_1, \cdots, v_q)$$

where each $a$ and each $\alpha$ is *algebraic* in all of its variables, as a product of functions of similar type.

In Section II, the existence of *irreducible* functions is proved. It is shown that

$$1 + xy + \phi(x)\psi(y),$$

where $\phi$ and $\psi$ are *transcendental*, cannot be the product of two functions of type (1), each of which, in reduced form, contains more than one term.

## I. Algebraic dependence

1. **Reduced form.** It is plain that if the $a$'s or $\alpha$'s in (1) are not linearly independent, we can write the function (1) with fewer than $n$ terms. We shall prove that when the $a$'s and $\alpha$'s are linearly independent, then $n$ is a minimum.

The linear independence of the $a$'s and $\alpha$'s being understood, suppose that (1) has another expression,

$$b_1(x)\beta_1(y) + \cdots + b_m(x)\beta_m(y). \tag{2}$$

We equate (1) and (2) and differentiate the resulting relation $n-1$ times with respect to $y$. Let $y_0$ be a point at which the wronskian of the $\alpha$'s does not vanish. Putting $y=y_0$ in the $n$ relations secured above, we solve for the $a$'s. We find each $a$ expressed as a linear combination of the $b$'s with constant coefficients. Hence we cannot have $m<n$, else the $a$'s would be linearly dependent. Thus $n$ is a minimum. We see also that, if we have two expressions of (1) in reduced form, the $a$'s and $\alpha$'s in either expression are linear combinations of those in the other.

2. **Algebraic dependence.** We shall prove the following theorem:

Theorem I. *Let*

$$A = a_1(x)\alpha_1(y) + \cdots + a_n(x)\alpha_n(y),$$

*with all a's analytic in some circle* $|x-\zeta|<\rho$ *and all b's analytic in some circle* $|y-\eta|<\sigma$. *Let* $A=BC$, *where*

$$B = b_1(x)\beta_1(y) + \cdots + b_r(x)\beta_r(y), \qquad C = c_1(x)\gamma_1(y) + \cdots + c_s(x)\gamma_s(y)$$

*with $b$'s and $c$'s all analytic in some area in $|x-\zeta|<\rho$ and $\beta$'s and $\gamma$'s all analytic in some area in $|y-\eta|<\sigma$. It is understood that $A$, $B$, $C$ are in reduced form. Then every $b_ic_j$ is algebraic in the $a$'s, and every $\beta_i\gamma_j$ is algebraic in the $\alpha$'s.*

That is, for every $b_ic_j$, there exists, identically in $x$, a relation

$$P_m b_i^m c_j^m + \cdots + P_1 b_i c_j + P_0 = 0,$$

where the $P$'s are polynomials in the $a$'s with constant coefficients, not all $P$'s being identically zero. A similar relation holds for every $\beta_i\gamma_j$.

We shall prove that part of our statement which refers to the $a$'s, $b$'s and $c$'s. The functions of $y$ are treated similarly.

The product of $B$ by $C$ is of the form

$$\sum b_ic_j\beta_i\gamma_j \qquad (i = 1, \cdots, r; j = 1, \cdots, s). \tag{3}$$

Let

$$\beta_{i_1}\gamma_{j_1}, \cdots, \beta_{i_p}\gamma_{j_p} \tag{4}$$

be linearly independent, and such that every $\beta_i\gamma_j$ is a linear combination of them with constant coefficients. Then

$$BC = S_1\beta_{i_1}\gamma_{j_1} + \cdots + S_p\beta_{i_p}\gamma_{j_p} \tag{5}$$

where each $S$ is a linear combination of the functions $b_ic_j$, with constant coefficients.

The $S$'s may be linearly dependent. Suppose, changing subscripts, if necessary, that

$$S_1, \cdots, S_q, \quad q \leqq p,$$

are linearly independent, and that every other $S_i$ is a linear combination of them with constant coefficients. If $q<p$, let, for $i>q$,

$$S_i = u_{i1}S_1 + \cdots + u_{iq}S_q, \tag{6}$$

with constant $u$'s. We have

$$BC = S_1\Sigma_1 + \cdots + S_q\Sigma_q, \tag{7}$$

where each $\Sigma$ is a linear combination of the functions in (4). The $\Sigma$'s are linearly independent, for, as each $\Sigma_k$ involves $\beta_{i_k}\gamma_{j_k}$, but no other $\beta_{i_k}\gamma_{j_k}$ with $k\leqq q$, a linear dependence among the $\Sigma$'s would imply one among the functions of (4).

Thus $q=n$, and (7) is a representation of $A=BC$ in reduced form. We

execute linear transformations upon the $a$'s and $\alpha$'s in $A$, in such a way as to have, for the new $a$'s and $\alpha$'s,

$$S_i = a_i \qquad (i = 1, \cdots, n), \tag{8}$$

$$\Sigma_i = \alpha_i \qquad (i = 1, \cdots, n). \tag{9}$$

Evidently, if the products $b_ic_j$ are algebraic in the new $a$'s, they are also algebraic in the original $a$'s.

We have, in (6) and (8), a system of linear equations which are satisfied by the $b_ic_j$'s. Now every $b_ic_j$ can be expressed in terms of the $s+r-1$ products

$$b_1c_1, b_1c_2, \cdots, b_1c_s; b_2c_s, \cdots, b_rc_s. \tag{10}$$

We have

$$b_ic_j = \frac{(b_1c_j)(b_ic_s)}{b_1c_s}. \tag{11}$$

If a $b_ic_j$ is not in (10), we replace it in (6) and (8) by its expression (11). The equations (6) and (8) go over into equations in the functions of (10). Multiplying those of the latter equations in which $b_1c_s$ appears in a denominator by $b_1c_s$, we find a system of $(p-q)+n=p$ algebraic equations

$$E_i = 0 \qquad (i = 1, \cdots, p), \tag{12}$$

which are satisfied by the products (10). In the first $p-q=p-n$ of the equations (12) the coefficients are constants. In each of the last $n$ equations, an $a$ appears among the coefficients.

The remainder of our proof will be concerned with showing that (12) determines the products (10) as algebraic combinations of the $a$'s.

Let us, for the moment, regard the $b_ic_j$'s in (12) as any solutions of (12), rather than as the products of the fixed $b$'s and $c$'s in $B$ and $C$. That is, momentarily, $b_ic_j$ is not being regarded as a product. Consider any solution of (12) with $b_1c_s \neq 0$. Let a function $b_1 \neq 0$ be chosen arbitrarily. We can now determine $c_1, \cdots, c_s$ in such a way that $b_1c_i$ is the product of $b_1$ by $c_i$, $i=1, \cdots, s$, and, having $c_s$, we can determine $b_2, \cdots, b_r$ so that $b_ic_s$ is the product of $b_i$ by $c_s$, $i=2, \cdots, s$. Then the products $b_ic_j$, $i=1, \cdots, r$; $j=1, \cdots, s$, will satisfy (6) and (8). If, now, we consider the product

$$(b_1\beta_1 + \cdots + b_r\beta_r)(c_1\gamma_1 + \cdots + c_s\gamma_s),$$

where the $b$'s and $c$'s are those just determined and where the $\beta$'s and $\gamma$'s are the fixed functions appearing above, we see first that this product equals the second member of (5), with the original $b$'s and $c$'s replaced by the new ones, then that it equals the second member of (7), with the same substitu-

## SUGGESTIONS TO AUTHORS

**Much needless expense and many errors can be avoided. The editors of several mathematical journals have agreed upon the following suggestions.**

1. Typewrite words and the very simplest formulas only.
2. *Do not* try to typewrite any complex formulas. Write them.
3. Keep a copy, and send the editors two copies, if you can.
4. *Do not* underline any symbols or any formulas.
5. Underline theorems with blue pencil (avoid ink).
6. Follow our recent styles in abbreviations, footnotes, etc.
7. Write carefully the (often misunderstood) capitals C K P S V W X Z.
8. Write $\epsilon$, not $\varepsilon$. Write very carefully $\gamma\ \eta\ \kappa\ \lambda\ \nu\ \tau\ \upsilon\ \chi\ \omega$.
9. Among Greek capitals, use only $\Gamma\ \Delta\ \Theta\ \Lambda\ \Xi\ \Pi\ \Sigma\ \Phi\ \Psi\ \Omega$.
10. Punctuate carefully, especially in formulas; thus: $1, 2, \cdots, n$.
11. Use the solidus (/) to avoid fractions in solid lines.
12. Use fractional exponents to avoid root signs everywhere.
13. Use extra symbols to avoid complicated exponents.
14. In typewritten formulas, $1$ means "one"; to indicate "ell" in formulas, backspace and overprint /; thus: $l$. Similarly, means "zero"; to indicate "cap O," backspace and overprint period; thus: $O$.
15. Avoid a dash over a letter, except for those shown below.
16. Some samples of unusual types available on monotype machines follow. A more complete list of all such types will be sent on request.

---

Light Face Greek—$\alpha\ \beta\ \gamma \cdots$ (all) A B $\Gamma \cdots$ (all).

★ Light Greek Superiors—$^{\Delta}$ and $^{\alpha\ \beta\ \gamma \cdots}$ (all except $^{o}$).

★ Light Greek Inferiors—$_{\Delta\ \Lambda\ \Sigma\ \Omega}$ and $_{\alpha\ \beta\ \gamma \cdots}$ (all except $_{o}$).

* Boldface Greek—$\boldsymbol{\alpha\ \beta\ \gamma\ \delta\ \varepsilon\ \zeta\ \eta\ \theta\ \varkappa\ \lambda\ \mu\ \nu\ \xi\ \pi\ \varrho\ \sigma\ \omega}$ and $\boldsymbol{\Omega}$.

* Lightface German—$\mathfrak{a\ b\ c\ d\ e\ f\ p\ q}\ \mathfrak{A\ B\ C\ D\ E\ F\ G\ H\ J\ K\ L\ M\ P\ Q\ R\ S\ T\ U\ V\ W\ X\ Y\ Z}$.

* Boldface German—$\mathfrak{b}\ \mathfrak{A\ B\ D}$.

Script (special font) $\mathcal{A\ B\ C} \cdots$ (all). No lower case manufactured.

* Hebrew— א נ ב troublesome to handle.

★ Dashed Italics—$\overline{A}\ \bar{a}\ \overline{B}\ \bar{b}\ \overline{C}\ \bar{c}\ \overline{E}\ \bar{e}\ \overline{F}\ \bar{f}\ \overline{G}\ \bar{g}\ \overline{H}\ \bar{l}\ \bar{\imath}\ \bar{\jmath}\ \overline{K}\ \bar{k}\ \overline{M}\ \bar{m}\ \overline{N}\ \bar{n}\ \bar{o}\ \overline{P}\ \bar{p}\ \bar{q}\ \overline{R}\ \bar{r}\ \bar{s}\ \bar{t}\ \bar{u}\ \bar{v}\ \overline{X}\ \bar{x}\ \overline{Y}\ \bar{y}\ \overline{Z}\ \bar{z}$.

★ Tilda Italics—$\tilde{A}\ \tilde{a}\ \tilde{e}\ \tilde{N}\ \tilde{n}\ \tilde{O}\ \tilde{r}\ \tilde{u}\ \tilde{y}$.

★ Tilda Greek—$\tilde{\alpha}\ \tilde{\epsilon}\ \tilde{\eta}\ \tilde{\omega}\ \tilde{\varpi}$.

★ Dashed Greek—$\bar{\alpha}\ \bar{\beta}\ \bar{\gamma}\ \bar{\delta}\ \bar{\eta}\ \bar{\theta}\ \bar{\mu}\ \bar{\nu}\ \bar{\rho}\ \bar{\omega}\ \bar{\Gamma}$.

★ Dotted Italic—$\dot{a}\ \ddot{a}\ \dot{e}\ \ddot{e}\ \dot{g}\ \ddot{\imath}\ \dot{m}\ \dot{n}\ \dot{q}\ \dot{r}\ \ddot{r}\ \dot{u}\ \ddot{u}\ \dot{x}\ \ddot{x}\ \dot{y}\ \ddot{y}\ \dot{z}\ \ddot{z}$.

★ Dotted Greek—$\dot{\eta}\ \ddot{\eta}\ \dot{\theta}\ \ddot{\theta}\ \dot{\xi}\ \dot{\psi}\ \ddot{\psi}\ \tilde{\omega}$ (single dotted $\zeta\ \phi\ \delta\ \beta\ \gamma$; double dotted $\gamma$ readily available).

---

* Additional characters readily available at small cost.

★ Matrices for additional characters are made upon special orders and necessitate a delay of from four to eight weeks and average expense of $4.50 per matrix.

tions, finally that it equals the fixed function $A$. Hence any solution of (12) with $b_1c_s \not\equiv 0$ will give a representation of $A$ as a product of two functions of type (1).

We are going to discuss the totality of the solutions of (12), using the theory of systems of algebraic equations as presented by van der Waerden in his paper *Zur Nullstellentheorie der Polynomideale.**

We consider the ideal, with coefficients in the field of the $a$'s and all constants, of which the functions $E_i$ in (12) are the basis. We represent this ideal as the least common multiple of a finite number of primary ideals. There is one of these primary ideals whose polynomials all vanish for the fixed $b_ic_j$'s in (3). We take this primary ideal, and form the prime ideal associated with it, that is, the set of polynomials in the field of the $a$'s and constants, which are such that some power of every one of them is in the primary ideal.

Let $F_1, \cdots, F_l$ be a basis for the prime ideal. Naturally every solution of

$$F_1 = 0, \cdots, F_l = 0 \tag{13}$$

satisfies (12).

If (13) has only a finite number of solutions, then every $b_ic_j$ in any of its solutions is algebraic in the $a$'s.

Thus our theorem will be proved if we can show that (13) has only a finite number of solutions. Suppose that (13) has an infinite number of solutions. Then certain of the quantities (10) can be chosen arbitrarily, and (13) determines the remaining quantities as algebraic combinations of these.

What we have to keep in mind, in a function-theoretic way, is this. If the arbitrary quantities (10) are represented by $z_1, \cdots, z_\mu$, there exist $s+r-1-\mu$ functions of $x, z_1, \cdots, z_\mu$, all analytic at a point $(\chi, \zeta_1, \cdots, \zeta_\mu)$ where $\chi$ is some point at which the $a$'s are analytic,† such that, if the $z$'s are taken as functions of $x$ analytic at $\chi$ and assuming, at $\chi$, values close to the $\zeta$'s, we get a set of $s+r-1$ analytic functions of $x$ which satisfy (13) (also (12)). For each $x$, the $s+r-1-\mu$ functions above are algebraic in the $z$'s.

If $b_1c_s$ does not correspond to one of the arbitrary $z$'s, then the function of $x$ and the $z$'s, found for $b_1c_s$, will not be identically zero, for (13) contains a special solution, that of the original quantities (10), in which $b_1c_s$ is not zero.

We shall keep $z_2, \cdots, z_\mu$ fixed, at constant values close to their $\zeta$'s. We shall keep $z_1$ indeterminate, allowing it to range over the neighborhood

* Mathematische Annalen, vol. 96 (1927), p. 196.

† For $\zeta_1, \cdots, \zeta_\mu$ we can take any values not situated on certain singular manifolds.

of $\zeta_1$. Then the quantities (10) go over into functions analytic in $x$ and algebraic in $z_1$. If $b_1c_s$ is among $z_2, \cdots, z_\mu$, we take care to give it a constant value distinct from zero. If $b_1c_s$ is not among $z_2, \cdots, z_\mu$, we choose $z_2, \cdots, z_\mu$ so that $b_1c_s$, as a function of $x$ and $z_1$, is not identically zero.

We have now a one-parameter family of representations of $A$ in the form $BC$.* For $b_1$, we take any function of $x$, analytic and not zero at $\chi$. This determines the remaining $b$'s and $c$'s as functions of $x$ and $z_1$. We write

$$\begin{aligned} B &= b_1(x)\beta_1 + b_2(x,z_1)\beta_2 + \cdots + b_r(x,z_1)\beta_r, \\ C &= c_1(x,z_1)\gamma_1 + \cdots + c_s(x,z_1)\gamma_s. \end{aligned} \tag{14}$$

We shall prove that either two of the $b$'s in (14), or else two of the $c$'s, have a ratio which is not a function of $x$ alone.† Suppose that the contrary is true. Then, as $b_1$ is independent of $z_1$, every $b$ is independent of $z_1$. Then the $c$'s cannot all be independent of $z_1$, for some $b_ic_j$ equals $z_1$. Hence $C$ is the product of $c_1(x, z_1)$ by a function of $x$ and $y$ alone. A similar fact is true of $A$. Then $A$, like $c_1(x, z_1)$, must vary with $z_1$. This contradiction proves our statement.

Hence there must be some value $x'$ of $x$ which makes neither $B$ nor $C$ vanish identically in $y$ and $z_1$, and for which at least two of the $b$'s, or two of the $c$'s, become functions of $z_1$ which are linearly independent. We write

$$A(x',y) = B(x',y,z_1)C(x',y,z_1), \tag{15}$$

where, expressing $B$ and $C$ in reduced form as sums of products of functions of $y$ and functions of $z_1$, we have

$$B(x',y,z_1) = \delta_1(y)d_1(z_1) + \cdots + \delta_t(y)d_t(z_1), \tag{16}$$

$$C(x',y,z_1) = \epsilon_1(y)e_1(z_1) + \cdots + \epsilon_u(y)e_u(z_1). \tag{17}$$

$A(x', y)$ is not zero for every $y$. In (16), (17), we must have either $t>1$ or $u>1$, for since the $\beta$'s are linearly independent, and also the $\gamma$'s, the elimination of the linear dependences among the $b$'s and $c$'s of (14), with $x=x'$, is sufficient for the conversion of $B$ and $C$ into reduced form.

We let

$$\theta_i = \frac{\delta_i}{A(x',y)} \qquad (i = 1, \cdots, t).$$

Then

$$[\theta_1d_1(z_1) + \cdots + \theta_td_t(z_1)][\epsilon_1e_1(z_1) + \cdots + \epsilon_ue_u(z_1)] = 1 \tag{18}$$

---

* $B$ and $C$ may not be in reduced form in these representations.

† What follows also proves that either $r>1$ or $s>1$.

for every $y$ and $z_1$. We represent, from now on, the first and second factors of the first member of (18) by $P$ and $Q$ respectively.

Since the $d$'s and $e$'s are algebraic in $z_1$, there is a Riemann surface on which they are all uniform. Let $R$ be such a surface. Let $U$ be an area in the $y$-plane in which the $\theta$'s and $\epsilon$'s are all analytic. For every $y$ in $U$, $P$ and $Q$ are algebraic functions of $z_1$, uniform on $R$.

Suppose that some $d$ has a pole at a point on $R$, corresponding to the value $w$ of $z_1$. First, let $w$ be finite. Of all $d$'s which have poles at this point, let $d_{i_1}, \cdots, d_{i_m}$ be those the order of whose poles is a maximum, and let their developments at the pole all begin with terms in $(z_1-w)^h$ with $h<0$. Then, for any $y$ in $U$, $P$ has a development at the point under consideration beginning with a term in $(z_1-w)^h$ of coefficient

$$v_{i_1}\theta_{i_1} + \cdots + v_{i_m}\theta_{i_m},$$

where the $v$'s are constants. As the $\theta$'s are linearly independent, there is an area $V$, in $U$, in which the coefficient does not vanish. For any $y$ in $V$, $P$ will have a pole at the point under consideration. A similar result holds for $w=\infty$.

Evidently, then, we can find an area $W$, in $U$, such that, for $y$ in $W$, $P$ has a pole where and only where one of the $d$'s has a pole, and $Q$ has a pole where and only where one of the $e$'s has a pole. We may furthermore suppose that the order of any of the poles of $P$ or of $Q$ is independent of $y$ (in $W$).

As either $t>1$ or $u>1$, we suppose, without loss of generality, that $t>1$. By (18), $P$ has a zero where and only where $Q$ has a pole. Hence the zeros of $P$, as $y$ ranges over $W$, are fixed points, and the order of any zero of $P$ is independent of $y$.

Consider any two values of $y$ in $W$. For each of these values, $P$ becomes an algebraic function of $z_1$. The two algebraic functions, having the same zeros and the same poles, have a constant ratio. That is, $P$ is a function of $y$ multiplied by a function of $z_1$. This contradiction of the fact that $P$, in reduced form, has $t>1$ terms, shows that (13) has only a finite number of solutions. Theorem I is proved.

3. **Functions with one term.** We prove the following theorem:

THEOREM II. *Unity cannot be the product of two functions* (1), *either of which, in reduced form, contains more than one term.*

Let $1=BC$, with $B$ and $C$ in reduced form. Arrange so that $b_1=1, \beta_1=1$. Then every $b$ and $c$ will be an algebraic combination of 1, and hence will be

a constant. Thus, if there were two $b$'s, they would be linearly dependent. This proves our theorem.

4. **Reduction of problem.** Let $A=BC$, as in Theorem I. Suppose that $b_1=1$, $\beta_1=1$, so that the $b$'s and $c$'s are algebraic in the $a$'s, the $\beta$'s and $\gamma$'s algebraic in the $\alpha$'s. We suppose that $n>1$.

Let $a_1, \cdots, a_p$ be such that no algebraic relation exists among them, but that every other $a$ is an algebraic combination of them. Let $\alpha_1, \cdots, \alpha_q$ be a similar set of $\alpha$'s. Then the $b$'s and $c$'s are all algebraic in $a_1, \cdots, a_p$, the $\beta$'s and $\gamma$'s algebraic in $\alpha_1, \cdots, \alpha_q$.

Let $w$ be any function of $x$ which is algebraic in $a_1, \cdots, a_p$. Let

$$P_0 w^m + \cdots + P_m = 0, \tag{19}$$

where every $P$ is a polynomial in $a_1, \cdots, a_p$. We suppose (19) to be irreducible in the field of $a_1, \cdots, a_p$ and all constants.

We replace $a_1, \cdots, a_p$ in the $P$'s by variables $u_1, \cdots, u_p$. Then $w$ becomes an algebraic function of the $u$'s. The singularities of this algebraic function are characterized by algebraic relations among the $u$'s. Hence, because $a_1, \cdots, a_p$ are algebraically independent, not all values of the $u$'s which are values of $a_1, \cdots, a_p$ can be singularities of the algebraic function. Thus there is a branch of the algebraic function which is analytic for values of the $u$'s which are values of the $a$'s, and which reduces to the original function $w$ of $x$ when the $u$'s are replaced by the $a$'s.

Thus, to the relation $A=BC$, there corresponds a relation

$$\begin{aligned} &a_1(u_1, \cdots, u_p)\alpha_1(v_1, \cdots, v_q) + \cdots + a_n(u_1, \cdots, u_p)\alpha_n(v_1, \cdots, v_q) \\ &= [b_1(u_1, \cdots, u_p)\beta_1(v_1, \cdots, v_q) + \cdots + b_r(u_1, \cdots, u_p)\beta_r(v_1, \cdots, v_q)] \\ &\cdot [c_1(u_1, \cdots, u_p)\gamma_1(v_1, \cdots, v_q) + \cdots + c_s(u_1, \cdots, u_p)\gamma_s(v_1, \cdots, v_q)], \end{aligned} \tag{20}$$

all functions being algebraic in their arguments. The relation holds when each $u_i$ is replaced by $a_i(x)$, each $v_i$ by $b_i(x)$. The functions of the $u$'s in (20) are all analytic for certain values of the $u$'s which are values of $a_1(x), \cdots, a_p(x)$. A similar statement holds for the functions of the $v$'s.

We say that (20) is an *identity* in the $u$'s and $v$'s.

If the $u$'s are held fixed at values assumed by $a_1(x), \cdots, a_p(x)$ at any $x$, then (20) becomes an identity in the $v$'s. If it did not, there would exist a non-identical algebraic relation among $\beta_1(y), \cdots, \beta_q(y)$. This means that (20) holds for arbitrary $v$'s, if only the $u$'s are replaced by the $a$'s. Then, for any set at all of $v$'s, (20) must be an identity in the $u$'s. Thus (20) is an identity in the $u$'s and $v$'s.

Conversely, suppose that the first member of (20) has, identically, such

a factorization as is displayed in the second member. As we saw above, no $b$ or $c$ can have a singularity for all values of the $u$'s which are values of $a_1(x), \cdots, a_p(x)$. A similar fact holds for the $\beta$'s and $\gamma$'s in (20). Hence every factorization of the first member of (20) gives a factorization $A=BC$. Furthermore, if the $b$'s, for instance, in (20), are linearly independent, the functions of $x$ obtained from them will be linearly independent.

Thus there is a perfect equivalence between the problem of finding all factorizations of the function $A$, and the problem of determining all representations of the algebraic function

$$a_1(u_1, \cdots, u_p)\alpha_1(v_1, \cdots, v_q) + \cdots + a_n(u_1, \cdots, u_p)\alpha_n(v_1, \cdots, v_q)$$

as a product of similar functions.

## II. Irreducibility

**5. Statement of result.** We shall prove the following theorem:

Theorem III. *A function*

$$1 + xy + \phi(x)\psi(y),$$

*with $\phi$ and $\psi$ transcendental, cannot be the product of two functions of type* (1), *each of which, in reduced form, has more than one term.*

In virtue of §4, we have to prove that there exists no relation

$$1 + xy + uv = \left[\sum_{i=1}^{r} a_i(x, u)\alpha_i(y, v)\right]\left[\sum_{i=1}^{s} b_i(x,u)\beta_i(y,v)\right], \tag{21}$$

with $r>1$, $s>1$, with every $a$, $\alpha$, $b$, $\beta$ algebraic in each of its variables and with each of the four sets of functions $a$, $\alpha$, $b$, $\beta$ linearly independent.

We assume a relation (21) to exist, and force a contradiction.

**6. First step.** We shall prove that there exists an expression for $1+xy+uv$

$$\begin{aligned} 1 + xy + uv = &\left[\sum_{i=1}^{p} c_i(x)\gamma_i(y) + \sum_{i=p+1}^{r} c_i(x,u)\gamma_i(y,v)\right] \\ &\cdot\left[\sum_{i=1}^{q} d_i(x)\delta_i(y) + \sum_{i=q+1}^{s} d_i(x,u)\delta_i(y,v)\right] \end{aligned} \tag{22}$$

where $c_1, \cdots, c_p; d_1, \cdots, d_q$, independent of $u$, and $\gamma_1, \cdots, \gamma_p; \delta_1, \cdots, \delta_q$, independent of $v$, are such that

$$1 + xy = (c_1\gamma_1 + \cdots + c_p\gamma_p)(d_1\delta_1 + \cdots + d_q\delta_q). \tag{23}$$

The $c$'s, $d$'s, $\gamma$'s and $\delta$'s in (22) are linearly independent sets of algebraic functions.

In (21), we arrange so that the $a$'s and $b$'s are supported by a single Riemann manifold in the space of $x$ and $u$, and, similarly, that all the $\alpha$'s and $\beta$'s have the same manifold. This amounts, for the $a$'s and $b$'s, for instance, to determining an algebraic function $P$ of $x$ and $u$ such that every $a$ and every $b$ is rational in $P$, $x$ and $u$.

It is possible, in an infinite number of ways, to take an area $M$ in the plane of $x$ such that, for $x$ in $M$ and for $u$ small, any branch of any $a$ or $b$ can be expanded in ascending integral or fractional powers of $u$, the coefficients being algebraic functions of $x$, analytic in $M$. In what follows we deal with such expansions of any properly associated set of branches of the $a$'s and $b$'s.

Let the lowest of all powers of $u$ which figure in the expansions of all the $a$'s be $u^g$. As the $\alpha$'s are linearly independent, the first factor in the second member of (21) will have an expansion beginning with $u^g$. Let the lowest power in the expansions of the $b$'s be $u^h$. Then, because $1+xy$ is present in the first member of (21), we must have $h+g=0$. We multiply the $a$'s by $u^{-g}$ and the $b$'s by $u^g$. We are thus able to, and do, assume that the developments of the $a$'s and $b$'s contain no negative powers of $u$, and that some $a$, as well as some $b$, begins with a term in $u^0$. The multiplications by $u^g$ and $u^{-g}$ require us to place the $a$'s and $b$'s on a manifold on which $u^g$ is uniform, as well as $P$, above.

Suppose that $a_1, \cdots, a_p$ begin with terms $A_i(x)u^0$, $i=1, \cdots, p$, that $b_1, \cdots, b_q$ begin with $B_i(x)u^0$, $i=1, \cdots, q$, and that the other $a$'s and $b$'s begin with higher powers of $u$. It is understood that no $A_i$ or $B_i$ is zero for every $x$. Equating the terms in $u^0$ in both members of (21), we have

$$1 + xy = (A_1\alpha_1 + \cdots + A_p\alpha_p)(B_1\beta_1 + \cdots + B_q\beta_q). \tag{24}$$

It is permissible to assume that the $A$'s, and likewise the $B$'s, are linearly independent. For instance, if $A_1, \cdots, A_j (j<p)$, are linearly independent, and the other $A$'s linearly dependent upon them, we subtract from $a_{j+1}, \cdots, a_p$ linear combinations of $a_1, \cdots, a_j$, with constant coefficients, in such a way as to get a new set $a_{j+1}, \cdots, a_p$, whose developments begin with positive powers of $u$. This, of course, would require a suitable change of $\alpha_1, \cdots, \alpha_j$. We assume, then, that the $A$'s and $B$'s are linearly independent.

We are going to prove that the ratio of any two $\alpha$'s, if $p>1$, or of any two $\beta$'s, if $q>1$, in (24), is independent of $v$. Assuming the contrary, let us suppose, as we may, that $p>1$ and that $\alpha_2/\alpha_1$ is not independent of $v$. Let $y_0$ be a value of $y$ for which $\alpha_2/\alpha_1$ becomes a non-constant function of $v$. We put $y=y_0$ in (24) and have

$$1 + xy_0 = [A_1\alpha_1(y_0,v) + \cdots + A_p\alpha_p(y_0,v)] \cdot [B_1\beta_1(y_0,v) + \cdots + B_q\beta_q(y_0,v)]. \tag{25}$$

Now (25) gives a representation of $1+xy_0$ as a product of functions of $x$ and $v$ of type (1), one of the functions of $x$ and $v$ consisting of at least two terms. This contradicts Theorem II. Thus any two $\alpha$'s, or any two $\beta$'s, in (24), have a ratio independent of $v$.

We see also, from (24), that $1+xy$ equals $\alpha_1\beta_1$ multiplied by a function independent of $v$, so that $\alpha_1\beta_1$ is independent of $v$. In (21), we divide the first factor, and multiply the second factor, by $\alpha_1$. We have then*

$$1 + xy + uv = \left[\sum_{i=1}^{p} a_i\gamma_i(y) + \sum_{i=p+1}^{r} a_i\gamma_i(y,v)\right] \cdot \left[\sum_{i=1}^{q} b_i\delta_i(y) + \sum_{i=q+1}^{s} b_i\delta_i(y,v)\right] \tag{26}$$

with algebraic, and linearly independent, $\gamma$'s and $\delta$'s.

We take now an area $N$ in the plane of $y$ such that all of the branches of the $\gamma$'s and $\delta$'s have, for $v$ small, expansions in ascending powers of $v$, with coefficients algebraic in $y$ and analytic throughout $N$. In what follows, we work with any properly associated set of branches of the $\gamma$'s and $\delta$'s.

If $v^g$ is the lowest power of $v$ in the developments of the $\gamma$'s, then $v^{-g}$ must be the lowest power in the $\delta$'s, since the first member of (26) contains $1+xy$. But $g \leqq 0$, since $\gamma_1(y) = \gamma_1(y)v^0$. Similarly $-g \leqq 0$, so that $g=0$. That is, no $\gamma_i$ with $i>p$, or $\delta_i$ with $i>q$, starts with a negative power of $v$.

We shall show that it is permissible to assume that every $\gamma_i$ with $i>p$ and every $\delta_i$ with $i>q$ starts with a positive power of $v$. Suppose that $\gamma_{p+1}, \cdots, \gamma_{p+j}$ start with terms $C_i(y)v^0$, $i=p+1, \cdots, p+j$, that $\delta_{q+1}, \cdots, \delta_{q+k}$ start with terms $D_i(y)v^0$, while the $\gamma$'s and $\delta$'s with higher indices, if there are any, start with positive powers.

We shall prove that every $C_i$ is linearly dependent on $\gamma_1, \cdots, \gamma_p$, and every $D_i$ is linearly dependent on $\delta_1, \cdots, \delta_q$.

Suppose that

$$\gamma_1, \cdots, \gamma_p, \quad C_{p+1}, \cdots, C_{p+l} \qquad (l \leqq j) \tag{27}$$

are linearly independent, but that all other $C$'s, if there are any, are linearly dependent on the functions of (27). Similarly, suppose that

$$\delta_1, \cdots, \delta_q, D_{q+1}, \cdots, D_{q+m} \tag{28}$$

are linearly independent, but that all other $D$'s, if there are any, are linearly

* $\gamma_1(y) = 1$.

dependent on the functions of (28). Then, by subtracting suitable linear combinations of $\gamma_1, \cdots, \gamma_{p+l}$ from $\gamma_{p+l+1}, \cdots, \gamma_{p+j}$ (if the latter exist), we get a new set $\gamma_{p+l+1}, \cdots, \gamma_{p+j}$, which start with positive powers of $v$. This requires a modification of the first $p+l$ of the $a$'s, but it is to be observed that the first $p$ of the $a$'s, even if they should be changed, will start with zero powers of $u$, while $a_{p+1}, \cdots, a_{p+l}$ will start with positive powers. Similar remarks apply to the $\delta$'s and the $b$'s.

With these adjustments, we put $v=0$ in (26), and find

$$1 + xy = [a_1\gamma_1 + \cdots + a_p\gamma_p + a_{p+1}C_{p+1} + \cdots + a_{p+l}C_{p+l}] \cdot [b_1\delta_1 + \cdots + b_q\delta_q + b_{q+1}D_{q+1} + \cdots + b_{q+m}D_{q+m}]. \tag{29}$$

Referring to the discussion of (24), we see that, because the sets (27) and (28) are linearly independent, the ratio of any two $a$'s, or any two $b$'s, in (29), is independent of $u$. But this cannot be, for $a_{p+1}$ and $b_{q+1}$ start with higher powers of $u$ than do $a_1$ and $b_1$.

Thus, the $C$'s and $D$'s are, respectively, linear combinations of the $\gamma$'s and $\delta$'s. By subtracting suitable linear combinations of $\gamma_1, \cdots, \gamma_p$ from $\gamma_{p+1}, \cdots, \gamma_{p+j}$ and linear combinations of $\delta_1, \cdots, \delta_q$ from $\delta_{q+1}, \cdots, \delta_{q+k}$, we may arrange so that all $\gamma_i$ with $i>p$ and all $\delta_i$ with $i>q$ start with positive powers of $v$. As we have already seen, the modified $a_1, \cdots, a_p$; $b_1, \cdots, b_q$ will still start with zero powers of $u$.

We now put $v=0$ in (26) and find

$$1 + xy = (a_1\gamma_1 + \cdots + a_p\gamma_p)(b_1\delta_1 + \cdots + b_q\delta_q).$$

As above, it follows that the ratio of any two $a$'s, or any two $b$'s, is independent of $u$. Also, $a_1b_1$ is independent of $u$.

Dividing the first factor in the second member of (26) by $a_1$ and multiplying the second factor by $a_1$, we find a relation (22) in which $c_1=\gamma_1=1$. In this relation, $c_i$ with $i>p$ and $d_i$ with $i>q$ start with positive powers of $u$, whereas, $\gamma_i$ with $i>p$ and $\delta_i$ with $i>q$ start with positive powers of $v$.

7. **Second step.** We shall show the existence of a relation

$$(1 + xy + uv) = [1 + xy + a_3(x,u)\alpha_3(y,v) + \cdots + a_r(x,u)\alpha_r(y,v)] \cdot [1 + b_2(x,u)\beta_2(y,v) + \cdots + b_s(x,u)\beta_s(y,v)]. \tag{30}$$

In (30), the $r$ and $s$ may be either the $r$ and $s$ of (21), or the $r$ and $s$ of (21) interchanged. All functions are algebraic, and the four sets

$$1, x, a_3, \cdots, a_r; \quad 1, b_2, \cdots, b_s;$$
$$1, y, \alpha_3, \cdots, \alpha_r; \quad 1, \beta_2, \cdots, \beta_s$$

are linearly independent.

We start from the relation (22), which we shall write in the form (21). That is, in (21), we understand that certain terms are free of $u$ and $v$, and that in the remaining terms the $\alpha$'s and $\beta$'s have expansions which start with positive powers of $v$. The fact that the remaining $a$'s and $b$'s start with positive powers of $u$, we shall not make use of. We shall also, at times, refer directly to the representation (22).

We write $Y=xy$, $U=uv$, so that (21) becomes

$$(31)\quad 1+Y+U=\left[\sum_{i=1}^{r} a_i\left(x,\frac{U}{v}\right)\alpha_i\left(\frac{Y}{x},v\right)\right]\left[\sum_{i=1}^{s} b_i\left(x,\frac{U}{v}\right)\beta_i\left(\frac{Y}{x},v\right)\right].$$

It is our plan to expand all functions in the second member of (31) in ascending powers of $v$ about $v=0$. For the $\alpha$'s and $\beta$'s, we use the same branches that we used above, replacing $y$ by $Y/x$. As regards the $a$'s and $b$'s, we use any set of branches associated with each other in (21); that is, we develop such a set of branches about $u=\infty$, and replace $u$ by $U/v$. Each $a$ and $b$ will develop into ascending powers of $v/U$, with coefficients algebraic in $x$.

Suppose then that we have found a development of the second member of (31) in powers of $v$, with coefficients involving $Y$, $x$, $U$. The lowest power of $v$ in this development must be $v^0$, and the coefficient of $v^0$ must be $1+Y+U$. Hence $1+Y+U$ is the product of the coefficients of the lowest powers of $v$ in the two factors of the second member of (31).

If $v^g$ is the lowest power in the development of the first factor, then $v^{-g}$ is the lowest power in the second factor.

We shall prove that the coefficient of $v^0$ in the first factor is not zero. That coefficient is

$$(32)\qquad c_1(x)\gamma_1\left(\frac{Y}{x}\right)+\cdots+c_p(x)\gamma_p\left(\frac{Y}{x}\right)+\text{terms involving positive powers of } U.$$

The $c$'s and $\gamma$'s are those which appear in (22). The "terms involving positive powers of $U$" are those which come from terms in (31) in which the $\alpha$ starts with a positive power of $v$. Such a term must contain a positive power of $U$, because it is a product of a term in the $\alpha$ by one in the $a$. As the $\alpha$ term involves $v$ in a positive power, the $a$ term must involve $v$ in a negative power and hence $U$ in a positive power.

Of course, (32) cannot be zero for every $Y$, $x$, $U$. Similarly, the coefficient of $v^0$ in the second factor is not zero.

Hence $g\leqq 0$, $-g\leqq 0$, so that $g=0$.

Each term in (32) is a non-negative rational power of $U$ multiplied by a product $\rho(x)\sigma(Y/x)$, with $\rho$ and $\sigma$ algebraic. We write (32)

$$\sum_{i=1}^{h} U^{l_i}\rho_i(x)\sigma_i(Y/x), \tag{33}$$

where $h$ is some positive integer. The $l$'s in (33) are not necessarily distinct. We understand that no power of $U$ in (33) has a zero sum for its coefficients. That is, where a sum of coefficients is zero, we suppress the power of $U$. Similarly the coefficient of $v^0$ in the second factor in (31) may be written

$$\sum_{i=1}^{k} U^{m_i}\xi_i(x)\zeta_i(Y/x) \tag{34}$$

with every $m_i$ rational and non-negative and with every $\xi$ and $\zeta$ algebraic. No power of $U$ in (34) has a zero sum for its coefficients.

We have

$$1 + Y + U = \left(\sum_{i=1}^{h} U^{l_i}\rho_i(x)\sigma_i(Y/x)\right)\left(\sum_{i=1}^{k} U^{m_i}\xi_i(x)\zeta_i(Y/x)\right). \tag{35}$$

We now develop each $\rho\sigma$ and each $\xi\zeta$ in powers of $x$, for the neighborhood of $x=0$. The development in $x$ of the second member of (35) must start with $(1+Y+U)x^0$. Let $x^\varrho$ be the lowest power in the first factor, $x^{-\varrho}$ that in the second factor. Multiplying together the coefficients of these lowest powers, we get a relation

$$1 + Y + U = (\sum H_{ij}U^{l_i}Y^{\lambda_j})(\sum K_{ij}U^{m_i}Y^{\mu_j}), \tag{36}$$

with $l$'s and $m$'s which are also present in (35), with constant $H$'s and $K$'s, and with rational $\lambda$'s and $\mu$'s.

Let $t$ be a positive common denominator for the non-negative rational numbers $l$ and $m$ in (36). Let $\tau$ be a positive common denominator for the rational (not necessarily positive) $\lambda$'s and $\mu$'s. Let us replace $U$ by $U^t$, $Y$ by $Y^\tau$ in (36). From (36) it follows that if no $l_i$ or $m_i$ is unity, $1+Y^\tau+U^t$, considered as a polynomial in $U$, is reducible in the field of $Y$ and all constants. But because $1+Y^\tau$, considered as a function of $Y$, has simple zeros at the $\tau$th roots of $-1$, the equation $1+Y^\tau+U^t=0$ defines $U$ as a function of $Y$ of $t$ branches, so that $1+Y^\tau+U^t$ is irreducible.

We conclude that some $l_i$ or some $m_i$ in (36) equals unity. Suppose that some $l_i$ is unity. As such an $l_i$ is also present in (35), we see, since the highest power of $U$ in the first member of (35) is $U$, that no $m_i$ is positive. Thus every $m_i$ is zero. That is, the second factor of (35) is the

$$d_1(x)\delta_1(y) + \cdots + d_q(x)\delta_q(y)$$

of (22), with $y$ replaced by $Y/x$. Let the coefficient of $U$ in the first factor of (35) be denoted by $P$. Then

$$1 = P[d_1(x)\delta_1(Y/x) + \cdots + d_q(x)\delta_q(Y/x)].$$

From Theorem II, it follows that $q=1$. Turning to (23), we have

$$1 + xy = d_1(x)\delta_1(y)[c_1(x)\gamma_1(y) + \cdots + c_p(x)\gamma_p(y)].$$

We must therefore have $p=2$. Furthermore, multiplying the first factor of (22), and dividing the second factor of (22), by $d_1\delta_1$, we have $1+xy+uv$ exhibited in the form (30).

8. **Third step.** We shall prove that every $b$ in (30) is independent of $x$, and every $\beta$ independent of $y$.

Suppose that $b_2$, for instance, is not independent of $x$. Let $u_0$ be a value of $u$ such that $b_2(x, u_0)$ is not a constant, and such that no $b$ or $a$ has a singularity at $(x, u_0)$ for every $x$.

Then $b_2(x, u_0)$ must have a pole for some value of $x$. Suppose first that there is a pole for a finite value, $h$, of $x$. Of all functions $b_i(x, u_0)$ which have poles at $h$, let

$$b_{i_1}(x, u_0), \cdots, b_{i_p}(x, u_0)$$

be those whose poles at $h$ have a maximum order, their developments starting with terms $c_{i_m}(x-h)^g$, $g<0$. Because of the linear independence of the $\beta$'s, the second factor in (30), considered as a function of $x$, will, for $u=u_0$ and $y$ and $v$ general, have a pole at $h$ with a first term

$$(c_{i_1}\beta_{i_1} + \cdots + c_{i_p}\beta_{i_p})(x-h)^g.$$

Hence the first factor must have a zero when $u=u_0$, $x=h$. Then no $a(x, u_0)$ can have a pole at $h$ and

$$1 + hy + a_3(h, u_0)\alpha_3(y, v) + \cdots + a_r(h, u_0)\alpha_r(y, v) = 0 \tag{37}$$

for $y$ and $v$ general. But (37) expresses a linear dependence among $1, y, \alpha_3, \cdots, \alpha_r$. We conclude that $b_2(x, u_0)$ cannot have a pole for a finite value of $x$.

Suppose now that $b_2(x, u_0)$ has a pole at $\infty$. We see, as above, that, for $y$ and $v$ general, the second factor in (30) has a pole at $\infty$. Since $1+xy+uv$ has only a simple pole at $\infty$ (when $y\neq 0$), the development at $\infty$ of the first factor in (30) cannot start with a power of $x$ as high as the first power. This means that some of the $a_i(x, u_0)$ have poles at $\infty$ with developments which start with the first power of $x$, and that there is a relation of the type

$$y + c_{i_1}\alpha_{i_1}(y, v) + \cdots + c_{i_q}\alpha_{i_q}(y, v) = 0.$$

This completes the proof that every $b$ is independent of $x$. Similarly every $\beta$ is independent of $y$.

9. **Completion of proof.** We shall show that $s=1$ in (30). This means that, in (21), either $s=1$ or $r=1$. Theorem III will be proved.

Let $F(u, v)$ be used to represent the second factor in (30). We shall develop every $a$ in (30) in descending powers of $x$ about $x=\infty$. If the coefficient of a power $x^h$ in the development of $a_i$ is $c_i(u)$, the coefficient of $x^h$ in the development of the second member of (30) is*

$$[c_1(u)\alpha_1 + \cdots + c_r(u)\alpha_r]F(u,v).$$

As the $\alpha$'s are linearly independent, and as the first member of (30) involves only $x$ and $x^0$, we conclude that every $a$ is a linear integral function of $x$. Similarly, the $\alpha$'s are linear integral functions of $y$.

Then (30) becomes

$$1 + xy + uv = [xyA(u,v) + xB(u,v) + yC(u,v) + D(u,v)]F(u,v),$$

where $A, B, C, D, F$ are sums of products of functions of $u$ by functions of $v$.

Equating coefficients of $xy$, we have

$$1 = A(u,v)F(u,v),$$

so that, by Theorem II, $F(u,v)$, when written in reduced form, has only one term. As $F(u, v)$ is the second member of (30) we have $s=1$ in (30), and the proof of Theorem III is completed.

---

* We understand that $a_1=1$, $a_2=x$, $\alpha_1=1$, $\alpha_2=y$.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# NEW RESULTS IN THE THEORY OF NORMAL DIVISION ALGEBRAS*

BY

A. ADRIAN ALBERT

1. **Introduction.** In 1905 L. E. Dickson defined a set of normal division algebras of order $n^2$ which were based on cyclic $n$-ics and are called cyclic algebras.† No further division algebras were known until, in 1923, F. Cecioni constructed‡ algebras based on a *non-cyclic* abelian equation of degree four and which were apparently new division algebras. Cecioni made no attempt to show that the algebras that he had constructed were non-cyclic, that is, not equivalent to the much simpler cyclic algebras. He had however found a type of algebras possibly containing non-cyclic algebras.

The present paper begins by a consideration of the necessary and sufficient conditions that a given algebra $A$ of order sixteen of the Cecioni type be an associative division algebra. The associativity conditions are reduced to the question of finding the solutions in integers of

$$\gamma_3^2 - \gamma_4^2\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^2 - \gamma_4^2\sigma)$$

where $\rho$ and $\sigma$ are integers such that neither $\rho$, $\sigma$, nor $\sigma\rho$ is a rational square. This equation has been treated in great detail by R. G. Archibald.§ The conditions that $A$ be a division algebra are reduced by algebraic theorems to the conditions either (a) $G$ is a quadratic non-residue of $\sigma_1$, or (b) $\sigma$ is a quadratic non-residue of $G_1$, or (c) $-\sigma_1 G_1$ is a quadratic non-residue of $\pi$, where $\gamma_1^2 - \gamma_2^2\rho = \Gamma^2 G$, $\sigma = \sigma_1\pi$, the numbers $\Gamma$ and $G$ are integers such that $G$ is a product of distinct primes, and $\pi$ is the highest common factor of $G$ and $\sigma$. But the author has shown that all normal division algebras of order 16 are of the Cecioni type.‖ We have therefore constructed, in terms of the single condition given by (a), (b), and (c) and in terms of the integer solutions of a single diophantine equation, all normal division algebras of order sixteen over the field of all rational numbers. For the special case of cyclic algebras the solution of the equation is known and $G$ becomes $\rho$ so that, since all of the

* Presented to the Society, December 27, 1929; received by the editors in December, 1929.

† *Algebras and their Arithmetics*, p. 65.

‡ Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 209–254.

§ These Transactions, vol. 30 (1928), pp. 819–837.

‖ These Transactions, vol. 31 (1929), pp. 253–260.

quadratic non-residues of any number are known, we have constructed all cyclic algebras of order sixteen over $R$.

The problem of discovering whether or not there exist any non-cyclic normal division algebras of order sixteen presents itself. An algebraic necessary and sufficient condition that an algebra $A$ in sixteen units be cyclic is found and, by its use, *it is shown that all of the algebras constructed by Cecioni are cyclic*, and not new. An investigation is then made of the diophantine necessary and sufficient conditions that $A$ be non-cyclic and it is proved that $A$ is non-cyclic if and only if two quartic forms with coefficients polynomials in $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ are not null forms.

2. **Linear associative normal algebras of order sixteen.** We shall consider linear algebras over $R$, the field of all rational numbers. The algebras will be constructed with a quartic equation with rational coefficients and Galois group $G_4$ as a foundation. As is well known every such equation may be reduced by a rational Tschirnhausen transformation to an equation

$$\omega^4 + p\omega^2 + n^2 = 0 \qquad (p \text{ and } n \text{ in } R), \tag{1}$$

where $\omega$ is a scalar variable and $R$ is the field of all rational numbers. Suppose that $i$ satisfies equation (1). If we write

$$2i^2 + p = w, \quad i + \frac{n}{i} = v, \quad u = \frac{w}{v}$$

then

$$w^2 = p^2 - 4n^2, \quad v^2 = 2n - p, \quad u^2 = -(p + 2n),$$

so that the algebraic field $R(i)$ is evidently the direct product of two quadratic fields $R(u)$ and $R(v)$, and every quantity of $R(i)$ is uniquely expressible in the form

$$a(i) = \alpha_1 + \alpha_2 u + \alpha_3 v + \alpha_4 uv \qquad (\alpha, \cdots, \alpha_4 \text{ in } R).$$

It is well known that the converse of this proposition is true and that

LEMMA 1. *Every field $R(i)$ generated by a root of a quartic* (1) *with Galois group $G_4$ is a direct product of two quadratic fields*

$$R(u), \quad R(v) \tag{2}$$

*where*

$$u^2 = \rho, \quad v^2 = \sigma, \quad \rho = \rho_1\pi, \quad \sigma = \sigma_1\pi \tag{3}$$

*and $\rho_1, \sigma_1, \pi$ are each products of distinct rational prime integers such that no two of $\rho_1, \sigma_1, \pi$ have a factor in common. Conversely every such direct product of two quadratic fields defines a quartic field generated by a quantity $i$ satisfying a quartic with group $G_4$.*

Let us examine the quantity $i$ of the above lemma, where we assume that $u$ and $v$ are given. Suppose that we define a quantity $i=(u+1)v$. Then $i^2=(\rho+1+2u)\sigma$, $2\sigma u=i^2-(\rho+1)\sigma$ and

$$\begin{aligned} i^4 &= \sigma^2[(\rho+1)^2+4(\rho+1)u+4\rho] \\ &= \sigma^2[2/\rho+(\rho+1)^2]+2\sigma(\rho+1)[i^2-(\rho+1)\sigma] \\ &= 2\sigma(\rho+1)i^2-\sigma^2(\rho-1)^2 \end{aligned}$$

so that $i$ satisfies

$$\phi(\omega)\equiv\omega^4-2\sigma(\rho+1)\omega^2+\sigma^2(\rho-1)^2=0, \tag{4}$$

with roots $i$, $-i$, $\sigma(\rho-1)/i$, $-\sigma(\rho-1)/i$. But if we write

$$\theta_1=-i, \qquad \theta_2=\frac{-\sigma(\rho-1)}{i}, \quad \theta_3=-\theta_2, \tag{5}$$

then

$$\theta_2=v(1-u), \quad \theta_1=(u+1)(-v), \tag{6}$$

$$\theta_3=uv-v, \tag{7}$$

so that $\theta_1$ is obtained from $i$ by replacing $v$ by $-v$, $\theta_2$ from $i$ by replacing $u$ by $-u$, and $\theta_3$ from $i$ by replacing both $u$ by $-u$ and $v$ by $-v$. We may write any polynomial $a(i)$ of $R(i)$ in the form

$$a=\alpha_1+\alpha_2u+\alpha_3v+\alpha_4uv \qquad (\alpha_1,\cdots,\alpha_4 \text{ in } R) \tag{8}$$

and shall utilize the notations

$$a(-u)=\alpha_1-\alpha_2u+\alpha_3v-\alpha_4uv, \tag{9}$$

$$a(-v)=\alpha_1+\alpha_2u-\alpha_3v-\alpha_4uv, \tag{10}$$

$$a(-u,-v)=\alpha_1-\alpha_2u-\alpha_3v+\alpha_4uv. \tag{11}$$

Consider the algebra $A$ with sixteen units given by the basis

$$i^rj_s \qquad (r,s=0,1,2,3;\ j_3=j_1j_2;\ j_0=1), \tag{12}$$

and the multiplication table given by $\phi(i)=0$,

$$(a+bj_1)(c+dj_1)=ac+g_1bd(\theta_1)+[ad+bc(\theta_1)]j_1 \tag{13}$$

for any $a, b, c, d$ of $R(i)$,

$$(U+Vj_2)(X+Yj_2)=(UX+VY'g_2)+(UY+VX')j_2 \tag{14}$$

for any quantities $U$, $V$, $X$, $Y$ of the form

$$X = a + bj_1, \quad X' = a(\theta_2) + b(\theta_2)\alpha j_1, \tag{15}$$

where

$$g_1, \quad g_2, \quad \alpha \tag{16}$$

are fixed quantities of $R(i)$. Cecioni has shown that if we define $g_3$ by

$$g_3 = g_1 g_2 \alpha(\theta_1), \tag{17}$$

so that

$$\alpha = \frac{g_3(\theta_1)}{g_1(\theta_1)g_2(\theta_1)}, \tag{18}$$

then the algebra $A$ defined above is associative if and only if

$$g_1 = g_1(\theta_1), \quad g_2 = g_2(\theta_2), \quad g_3 = g_3(\theta_3) \tag{19}$$

and if

$$g_3 g_3(\theta_1) = g_1 g_1(\theta_2) g_2 g_2(\theta_1). \tag{20}$$

We may obviously choose as a new basis for $A$ the quantities

$$j_s, \quad uj_s, \quad vj_s, \quad uvj_s \qquad (s = 0,1,2,3) \tag{21}$$

and obtain the following properties:

$$j_1 a = a(-v)j_1, \quad j_2 a = a(-u)j_2, \quad j_3 a = a(-u,-v)j_3, \tag{22}$$

since

$$j_1 a(i) = a(\theta_1)j_1, \quad j_2 a(i) = a(\theta_2)j_2, \quad j_3 a(i) = a(\theta_3)j_3. \tag{23}$$

For, as we have seen, $\theta_1$ is obtained from $i$ by replacing $v$ by $-v$ and leaving $u$ unaltered, so that any polynomial in $-i$ is obtained from the same polynomial in $i$ by replacing $v$ in it by $-v$ and $u$ by $u$, that is we obtain $a(-v)$. The other equations of (23) are obtained by symmetry. If we write

$$g_1 = \gamma_1 + \gamma_2 u + (\delta_1 + \delta_2 u)v \qquad (\gamma_1,\gamma_2,\delta_1,\delta_2 \text{ in } R),$$

then $g_1 = g_1(\theta_1) = g_1(-v)$ so that, by the linear independence of $1, u, v, uv$ with respect to $R$, we have

$$(\delta_1 + \delta_2 u)v = -(\delta_1 + \delta_2 u)v = 0.$$

Hence by symmetry

$$g_1 = \gamma_1 + \gamma_2 u, \quad g_2 = \gamma_3 + \gamma_4 v, \quad g_3 = \gamma_5 + \gamma_6 uv, \tag{24}$$

and conversely any $g_i$ in the form (24) satisfy conditions (19). Now (20) becomes

$$\gamma_5^2 - \gamma_6^2 \sigma\rho = (\gamma_1^2 - \gamma_2^2 \rho)(\gamma_3^2 - \gamma_4^2 \sigma) \tag{25}$$

and all of the associative algebras given by (12), (13), (14), (15), (18), (24) will be found when we find all solutions of (25). This problem has been considered in great detail by R. G. Archibald (loc. cit.) and the conditions for the existence of solutions of (25) have been found. We shall leave the associativity conditions in terms of the solutions of this diophantine equation. Let us assume now that in all further work the $\gamma_1, \cdots, \gamma_6$ satisfy (25).

**Definition.** A linear algebra $A$ over $R$ is said to be a normal algebra if the only quantities of $A$ commutative with every quantity of $A$ are rational numbers.

The writers in the theory of normal division algebras have assumed the algebras they constructed normal algebras, without proof. The proof is usually easy to give and we shall give it for the case we are considering.

THEOREM 1. *The algebra $A$ defined by* (12), (13), (14), (15), (18), (24), (25) *is a normal algebra when* $g_3 \neq 0$, $g_2 \neq 0$, $g_1 \neq 0$.

For $A$ contains a sub-algebra $\Sigma$ whose quantities are composed of all quantities of the form

$$X = a + bj_1 \qquad (a \text{ and } b \text{ in } R(i))$$

of $A$ and every quantity of $A$ is expressible in the form

$$c = X + Yj_2 \qquad (X \text{ and } Y \text{ in } \Sigma).$$

Suppose that $c$ were in $A$ and were commutative with every quantity of $A$. Then in particular

$$cu = (X + Yj_2)u = u(X - Yj_2) = uc = u(X + Yj_2),$$

and since $u$ has an inverse $(1/\rho)u$ in $A$ we have

$$u^{-1}u(X - Yj_2) = X - Yj_2 = u^{-1}uc = X + Yj_2.$$

By the linearity of $A$ we obtain $Yj_2 = -Yj_2 = 0$, $c = X = a + bj_1$. By using $vc = cv = v(a - bj_1)$ and the fact that $v$ has an inverse $\sigma^{-1}v$ in $A$ we obtain $b = 0$, $c = a_1 + a_2v$ with $a_1$ and $a_2$ in $R(u)$. Now $j_1^2 = g_1 \neq 0$ so that, since $g_1$ is in $R(u)$ and has an inverse when it is not zero, $j_1$ has an inverse $g^{-1}j_1$. Then $cj_1 = j_1c$ shows that $a_1 + a_2v = a_1 - a_2v$ and $a_2 = 0$, $c = a_1 = \alpha_1 + \alpha_2U$, $\alpha_1$ and $\alpha_2$ in $R$. Finally $j_2$ has an inverse $g_2^{-1}j_2$ and $cj_2 = j_2c$ gives $\alpha_2 = 0$ which proves that $c$ is in $R$ as was desired.

3. **New necessary and sufficient conditions that $A$ be a division algebra.** We shall assume that $\rho$ and $\sigma$ are numbers satisfying the conditions of Lemma 1, that $\gamma_1, \cdots, \gamma_6$ satisfy (25), and that $g_1 \neq 0$, $g_2 \neq 0$, $g_3 \neq 0$. We seek to find what further restrictions it is necessary and sufficient to impose on the parameters in order that $A$ be a division algebra. We shall first find

a new sufficient condition that $\Sigma$, the sub-algebra of $A$ of order eight defined by its basis

$$i^r j_s \qquad (r = 0,1,2,3;\ s = 0,1),$$

be a division algebra. It is known* that a necessary and sufficient condition that $\Sigma$ be a division algebra is that

$$g_1 \neq cc(-v)$$

for any $c$ of $R(i)$. We shall prove

LEMMA 2. *Algebra $\Sigma$ is a division algebra if there exists no $a \neq 0$ in $R(i)$ for which*

$$aa(-v)g_1 = f_1 \text{ in } R. \tag{26}$$

For suppose that the hypothesis of Lemma 2 were satisfied and yet $\Sigma$ were not a division algebra. Using the known necessary and sufficient condition, there would exist a polynomial $c$ in $R(u, v)$ for which

$$g_1 = cc(-v).$$

If $c=0$ then $g_1=0$ contrary to hypothesis. Hence $c \neq 0$ and has an inverse in the quartic field $R(i)$. Write $c^{-1}=a$ and thus $[c(-v)]^{-1}=a(-v)$. Hence

$$aa(-v)g_1 = 1.$$

But this is contrary to our hypothesis that $aa(-v)\, g_1$ is not in $R$ for any $a$. Hence $\Sigma$ is a division algebra. Write

$$a_1 = \alpha_1 + \alpha_2 u, \quad a_2 = \alpha_3 + \alpha_4 u \qquad (\alpha_1, \cdots, \alpha_4 \text{ in } R)$$

and

$$a = a_1 + a_2 v.$$

Then

$$aa(-v) = a_1^2 - a_2^2\sigma = \alpha_1^2 + \alpha_2^2\rho - \sigma(\alpha_3^2 + \alpha_4^2\rho) + 2(\alpha_1\alpha_2 - \sigma\alpha_3\alpha_4)u,$$

so that

$$aa(-v)g_1 = f_1 + f_2 u \tag{27}$$

where

$$f_1 = \gamma_1[(\alpha_1^2 + \alpha_2^2\rho) - \sigma(\alpha_3^2 + \alpha_4^2\rho)] + 2\gamma_2\rho(\alpha_1\alpha_2 - \sigma\alpha_3\alpha_4), \tag{28}$$

$$f_2 = \gamma_2[(\alpha_1^2 + \alpha_2^2\rho) - \sigma(\alpha_3^2 + \alpha_4^2\rho)] + 2\gamma_1(\alpha_1\alpha_2 - \sigma\alpha_3\alpha_4). \tag{29}$$

* Cf. L. E. Dickson, *Algebren und ihre Zahlentheorie*, p. 64.

Hence there exists a polynomial for which $aa(-v)g_1$ is in $R$ if and only if the quaternary quadratic form $f_2$ is a null form. Now

$$\gamma_2 f_2 = (\alpha_1\gamma_2 + \gamma_1\alpha_2)^2 + (\gamma_2^2\rho - \gamma_1^2)\alpha_2^2 - \sigma(\gamma_2\alpha_3 + \gamma_1\alpha_4)^2 - \sigma(\gamma_2^2\rho - \gamma_1^2)\alpha_4^2$$

and if we write

$$\epsilon_1 = \alpha_1\gamma_2 + \gamma_1\alpha_2, \quad \epsilon_2 = \gamma_2\alpha_3 + \gamma_1\alpha_4, \quad \epsilon_3 = \alpha_2, \quad \epsilon_4 = \alpha_4,$$

then

$$\gamma_2 f_2 = (\epsilon_1^2 - \sigma\epsilon_2^2) - (\gamma_1^2 - \gamma_2^2\rho)(\epsilon_3^2 - \sigma\epsilon_4^2).$$

We say that $f_2$ is a null form if and only if $\gamma_2 f_2$ is a null form. For if $\gamma_2$ were zero then $g_1 = \gamma_1$ and by taking $a = 1 = \alpha_1$, $\alpha_2 = \alpha_3 = \alpha_4 = 0$ we show that $f_2$ is a null form. When $\gamma_2 \neq 0$ the form $f_2$ is zero if and only if the non-zero multiple $\gamma_2 f_2$ is zero. The variables in $\gamma_2 f_2$ are all zero when and only when the variables in $f_2$ are all zero and we have proved

LEMMA 3. *Algebra $\Sigma$ is a division algebra if the quaternary quadratic form*

$$\epsilon_1^2 - \sigma\epsilon_2^2 - (\gamma_1^2 - \gamma_2^2\rho)(\epsilon_3^2 - \sigma\epsilon_4^2) \tag{30}$$

*is not a null form. This is equivalent to stating that $R(i)$ contains no polynomial $a(i)$ for which $aa(-v)g_1$ is in $R$.*

We shall next find a sufficient condition that $A$ itself be a division algebra under the assumption that $\Sigma$ is one. It is known (loc. cit.) that $A$ is a division algebra when $\Sigma$ is one if and only if

$$g_2 \neq X'X$$

for any $X$ of $\Sigma$. We shall prove that

LEMMA 4. *Algebra $A$ is a division algebra when $\Sigma$ is one if there exist no polynomials $b \neq 0$ and $d \neq 0$ in $R(i)$ such that*

$$bb(-u)g_2 = f_3, \quad dd(-u, -v)g_3 = f_5$$

*with $f_3$ and $f_5$ in $R$.*

For suppose that the hypotheses of Lemma 4 were satisfied and yet $A$ were not a division algebra. Then there must exist a quantity $X$ in $\Sigma$ for which $g_2 = X'X$. We may write

$$X = b(-v) + dj_1$$

where $b$ and $d$ are in $R(i)$. Then

$$X' = b(-u, -v) + d(-u)\alpha j_1$$

and

$$X'X = [b(-u, -v)b(-v) + d(-u)d(-v)\alpha g_1] + [b(-u, -v)d + d(-u)b\alpha]j_1,$$

so that

$$g_2 = b(-v)b(-u, -v) + d(-v)d(-u)\alpha g_1. \tag{31}$$

Now we know that, from (18) and $g_1 = g_1(-v)$,

$$\alpha = \frac{g_3(-v)}{g_1 g_2(-v)},$$

so that if (31) is true then

$$g_2 g_2(-v) = b(-v)b(-u, -v)g_2(-v) + d(-v)d(-u)g_3(-v) \tag{32}$$

and by replacing $i$ in (32) by $\theta_1$ and hence $v$ by $-v$ we have

$$\gamma_3^2 - \gamma_4^2\sigma = bb(-u)g_2 + dd(-u, -v)g_3. \tag{33}$$

But it is easily shown that if

$$b = \beta_1 + \beta_2 v + (\beta_3 + \beta_4 v)u, \tag{34}$$

$$d = \delta_1 + \delta_2 uv + (\delta_3 + \delta_4 uv)u, \tag{35}$$

then

$$bb(-u)g_2 = f_3 + f_4 v, \quad dd(-u, -v)g_3 = f_5 + f_6 uv,$$

where

$$f_3 = \gamma_3[\beta_1^2 + \beta_2^2\sigma - \rho(\beta_3^2 + \beta_4^2\sigma)] + 2\gamma_4\sigma(\beta_1\beta_2 - \rho\beta_3\beta_4), \tag{36}$$

$$f_4 = \gamma_4[\beta_1^2 + \beta_2^2\sigma - \rho(\beta_3^2 + \beta_4^2\sigma)] + 2\gamma_3(\beta_1\beta_2 - \rho\beta_3\beta_4), \tag{37}$$

$$f_5 = \gamma_5[\delta_1^2 + \delta_2^2\sigma\rho - \rho(\delta_3^2 + \delta_4^2\sigma\rho)] + 2\gamma_6\sigma\rho(\delta_1\delta_2 - \rho\delta_3\delta_4), \tag{38}$$

$$f_6 = \gamma_6[\delta_1^2 + \delta_2^2\sigma\rho - \rho(\delta_3^2 + \delta_4^2\sigma\rho)] + 2\gamma_5(\delta_1\delta_2 - \rho\delta_3\delta_4). \tag{39}$$

Hence when (33) is true

$$\gamma_3^2 - \gamma_4^2\sigma = f_3 + f_4 v + f_5 + f_6 uv. \tag{40}$$

But $f_4 \neq 0$ unless $b = 0$ and hence also $f_3 = 0$, by our hypothesis. Similarly $f_6 \neq 0$ unless $b = f_3 = 0$. But (40) implies that $f_4 = f_6 = 0$ since 1, $u$, $v$, $uv$ are linearly independent with respect to $R$. We have thus secured a contradiction and Lemma 4 is true.

Exactly as we showed in the proof of Lemma 3 that the form (29) was a null form if and only if (30) was a null form, we may show here that the form (37) is a null form if and only if

$$\gamma_4 f_4 = \epsilon_1^2 - \epsilon_2^2\rho - (\gamma_3^2 - \gamma_4^2\sigma)(\epsilon_3^2 - \epsilon_4^2\rho) \tag{41}$$

is a null form and that the form (39) is a null form if and only if

$$\gamma_6 f_6 = \epsilon_1^2 - \epsilon_2^2\rho - (\gamma_5^2 - \gamma_6^2\sigma\rho)(\epsilon_3^2 - \epsilon_4^2\rho) \tag{42}$$

is a null form. Let us suppose that the form (41) is not a null form but that (42) is a null form. Then

$$\mu_1^2 - \mu_2^2\rho - (\gamma_5^2 - \gamma_6^2\sigma\rho)(\mu_3^2 - \mu_4^2\rho) = 0 \tag{43}$$

for $\mu_1, \cdots, \mu_4$ in $R$ and not all zero. Using the value $\gamma_5^2 - \gamma_6^2\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho) \cdot (\gamma_3^2 - \gamma_4^2\sigma)$ we have

$$\mu_1^2 - \mu_2^2\rho - (\gamma_3^2 - \gamma_4^2\sigma)(\gamma_1^2 - \gamma_2^2\rho)(\mu_3^2 - \mu_4^2\rho) = 0. \tag{44}$$

Let us assume that $\gamma_2\gamma_4 \neq 0$. If $\mu_1 = \mu_2 = 0$ then $\mu_3^2 - \mu_4^2\rho = 0$, and $\mu_3 = \mu_4 = 0$ since $\rho$ is not a rational square. Hence $\mu_1$ and $\mu_2$ are not both zero. Write

$$\epsilon_1 = \mu_1, \quad \epsilon_2 = \mu_2, \quad \epsilon_3 = \gamma_1\mu_3 + \gamma_2\mu_4\rho, \quad \epsilon_4 = \gamma_1\mu_4 + \gamma_2\mu_3, \tag{45}$$

so that $\epsilon_3^2 - \epsilon_4^2\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^2 - \gamma_4^2\sigma)$. Then

$$\epsilon_1^2 - \epsilon_2^2\rho - (\gamma_3^2 - \gamma_4^2\sigma)(\epsilon_3^2 - \epsilon_4^2\rho) = 0 \tag{46}$$

for $\epsilon_1, \cdots, \epsilon_4$ in $R$ and $\epsilon_1$ and $\epsilon_2$ not both zero. This contradicts the hypothesis that (41) is not a null form. Hence when (41) is not a null form and $\gamma_2\gamma_4 \neq 0$ the form (42) is not a null form.

In the form $f_2$ write $\alpha_3 = \alpha_6\rho$, $\alpha_4 = \alpha_5$. Then (29) becomes

$$f_2 = \gamma_2[\alpha_1^2 + \alpha_2^2\rho - \sigma\rho(\alpha_5^2 + \alpha_6^2\rho)] + 2\gamma_1(\alpha_1\alpha_2 - \sigma\rho\alpha_5\alpha_6), \tag{47}$$

whence

$$\gamma_2 f_2 = (\gamma_2\alpha_1 + \gamma_1\alpha_2)^2 - (\gamma_1^2 - \gamma_2^2\rho)\alpha_2^2 - \sigma\rho[(\gamma_2\alpha_5 + \gamma_1\alpha_6)^2 - (\gamma_1^2 - \gamma_2^2\rho)\alpha_6^2],$$

which is a null form if and only if

$$\epsilon_1^2 - \sigma\rho\epsilon_2^2 - (\gamma_1^2 - \gamma_2^2\rho)(\epsilon_3^2 - \sigma\rho\epsilon_4^2) \tag{48}$$

is a null form. Similarly write in (37) $\beta_3 \equiv \rho\beta_6$, $\beta_4 \equiv \beta_5$, and it becomes

$$f_4 = \gamma_4[(\beta_1^2 + \beta_2^2\sigma) - \sigma\rho(\beta_5^2 + \beta_6^2\sigma)] + 2\gamma_3(\beta_1\beta_2 - \sigma\rho\beta_3\beta_4),$$

which is a null form if and only if

$$\epsilon_1^2 - \sigma\rho\epsilon_2^2 - (\gamma_3^2 - \gamma_4^2\sigma)(\epsilon_3^2 - \sigma\rho\epsilon_4^2) \tag{49}$$

is a null form. Hence (30) is a null form if and only if (48) is a null form while (41) is a null form if and only if (49) is a null form. Suppose that (30) were not a null form and (41) were a null form. Then there would exist rational numbers $\mu_1, \cdots, \mu_4$ in $R$ and not all zero such that

$$\mu_1^2 - \sigma\rho\mu_2^2 - (\gamma_3^2 - \gamma_4^2\sigma)(\mu_3^2 - \mu_4^2\sigma\rho) = 0.$$

As before we may state that in particular $\mu_1$ and $\mu_2$ are not both zero. But $\gamma_5^2-\gamma_6^2\sigma\rho=(\gamma_1^2-\gamma_2^2\rho)(\gamma_3^2-\gamma_4^2\sigma)$ and thus $\gamma_3^2-\gamma_4^2\sigma=\lambda^2(\gamma_5^2-\gamma_6^2\sigma\rho)(\gamma_1^2-\gamma_2^2\rho)$ where $1/\lambda=\gamma_1^2-\gamma_2^2\rho\neq 0$ since $\gamma_2\neq 0$, $\rho$ is not a square. Then

$$\mu_1^2-\sigma\rho\mu_2^2-\lambda^2(\gamma_1^2-\gamma_2^2\rho)(\gamma_5^2-\gamma_6^2\sigma\rho)(\mu_3^2-\mu_4^2\sigma\rho)=0,$$

and, by letting $\epsilon_1=\mu_1$, $\epsilon_2=\mu_2$, $\epsilon_3=\lambda(\gamma_6\mu_3+\gamma_6\mu_4\sigma\rho)$, $\epsilon_4=\lambda(\gamma_5\mu_4+\gamma_6\mu_3)$, we have a contradiction of the fact that (30) and hence (48) is not a null form. We thus obtain

LEMMA 5. *Algebra $A$ is a division algebra when it is associative, when $\gamma_2\neq 0$, $\gamma_4\neq 0$, and when the form*

$$\epsilon_1^2-\sigma\epsilon_2^2-(\gamma_1^2-\gamma_2^2\rho)(\epsilon_3^2-\sigma\epsilon_4^2) \tag{30}$$

*in the four rational variables $\epsilon_1,\cdots,\epsilon_4$ is not a null form. This form is a null form if and only if there exists a polynomial $a$ in $R(i)$ such that*

$$(aj_1)^2=\mu,\quad \mu \text{ in } R. \tag{50}$$

We shall consider in detail the form (30). Write

$$e_1=\epsilon_1+\epsilon_2 v,\quad e_2=\epsilon_3+\epsilon_4 v\qquad(\epsilon_1,\cdots,\epsilon_4 \text{ in } R).$$

Then

$$e_1e_1(-v)=\epsilon_1^2-\epsilon_2^2\sigma,\quad e_2e_2(-v)=\epsilon_3^2-\epsilon_4^2\sigma,$$

and $\epsilon_1^2-\epsilon_2^2\sigma=0$ if and only if $\epsilon_1=\epsilon_2=0$, that is, if and only if $e_1=0$. Similarly $e_2e_2(-v)=0$ if and only if $e_2=0$. Hence if (30) is a null form there exist quantities $e_1$ and $e_2$ both not zero and both in $R(v)$, such that if $e_3=e_1/e_2=\mu_1+\mu_2 v$ then

$$\gamma_1^2-\gamma_2^2\rho=\mu_1^2-\mu_2^2\sigma$$

with $\mu_1$ and $\mu_2$ in $R$ and not both zero. We may make $\gamma_1,\cdots,\gamma_6$ integers by replacing $j_1$ and $j_2$ by integer multiples of these quantities. It follows that (30) is a null form if and only if the indefinite ternary quadratic form (with integer variables $\lambda_1$, $\lambda_2$, $\lambda_3$)

$$\lambda_1^2-\lambda_2^2\sigma-(\gamma_1^2-\gamma_2^2\rho)\lambda_3^2$$

is a null form. Thus $A$ is a division algebra when it is associative, when $\gamma_2\gamma_4\neq 0$, and when the above form is not a null form.

We have found a new sufficient condition that $A$ be a division algebra. We shall prove this condition also a necessary condition. Suppose that $A$ were an associative normal division algebra and that the form were not a null form so that $A$ contained a polynomial $a(i)$ of $R(i)$ such that if $j=aj_1$ then $j^2=\mu$ in $R$. But the algebra

$$(1,v,j,vj)$$

is a generalized quaternion sub-algebra of $A$ over $R$ and has the multiplication table $v^2=\sigma, j^2=\mu, jv=-vj$. This is known to be impossible when $A$ is a division algebra.* Similarly when $\gamma_2=0$ we may take $j=j_1$ and have a contradiction. When $\gamma_4=0$ we take the sub-algebra

$$(1, u, j_2, uj_2)$$

and show that $A$ is not a division algebra. We have proved

THEOREM 2. *The set of all linear associative normal division algebras in sixteen units over $R$ is obtained by letting $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ range over all rational integers such that*

(a) *$\rho$ is a product of distinct primes,*
(b) *$\sigma$ is a product of distinct primes,*
(c) *neither $\rho$, $\sigma$ nor $\sigma\rho$ is a rational square,*
(d) $\gamma_2 \neq 0$, $\gamma_4 \neq 0$,
(e) $\gamma_5^2 - \gamma_6^2\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^2 - \gamma_4^2\sigma)$,
(f) *the ternary quadratic form in the variables $\lambda_1, \lambda_2, \lambda_3$*

$$\lambda_1^2 - \lambda_2^2\sigma - (\gamma_1^2 - \gamma_2^2\rho)\lambda_3^2$$

*is not a null form.*

But all ternary quadratic null forms are known.† We write $\gamma_1^2 - \gamma_2^2\rho = \Gamma^2 G$, $G = G_1\pi$ and $\sigma = \sigma_1\pi$ where $\Gamma$, $G_1$, $\pi$, $\sigma_1$, $G$ are integers, $G$ is a product of distinct primes and $\pi$ is the positive greatest common divisor of $\sigma$ and $G$. Then the form (f) is known to be a null form if and only if

$$\pi\lambda_1^2 - \sigma\lambda_2^2 - G_1\lambda_3^2$$

is a null form and this is true if and only if

$$G \equiv E_1^2 \bmod \sigma_1, \qquad \sigma \equiv E_2^2 \bmod G_1, \qquad -\sigma_1 G_1 \equiv E_3^2 \bmod \pi.$$

Hence we have the alternative theorem

THEOREM 2′. *We may replace* (f) *of Theorem* 2 *by the statement that either $G$ is a quadratic non-residue of $\sigma_1$, $\sigma$ is a quadratic non-residue of $G_1$, or $-\sigma_1 G_1$ is a quadratic non-residue of $\pi$.*

**4. An algebraic necessary and sufficient condition that $A$ be a cyclic algebra.** Consider the form

$$Q = a_1x_1^2 + \cdots + a_5x_5^2, \tag{51}$$

where $a_1, \cdots, a_5$ are non-zero integers not all having the same sign. It is

* The author's paper, Annals of Mathematics, 1929, on *The structure of direct products*, etc.
† Cf. P. Bachmann, *Arithmetik der Quadratischen Formen*, Chapter 8.

well known that there exist integers $x_1, \cdots, x_6$ not all zero for which $Q=0$, that is that $Q$ is a null form. As an obvious consequence we have

LEMMA 6. *Consider the form*

$$Q = \delta_1\xi_1^2 + \cdots + \delta_6\xi_6^2 \tag{52}$$

*where $\delta_1, \cdots, \delta_6$ are non-zero rational integers not all having the same sign. Then there exist rational integers $\xi_1, \cdots, \xi_6$ not all zero for which $Q=0$.*

We shall consider the form in $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$

$$\begin{aligned} R = {} & \sigma(\Delta_1\alpha_1^2 + \tau\Delta_1\alpha_2^2 + 2\tau\alpha_1\alpha_2) + (\lambda\Delta_1 + \tau\mu)[\beta_1^2 + \beta_2^2\tau - \sigma(\beta_3^2 + \beta_4^2\tau)] \\ & + 2\tau(\mu\Delta_1 + \lambda)(\beta_1\beta_2 - \sigma\beta_3\beta_4) \end{aligned} \tag{53}$$

where

$$\lambda,\ \sigma \neq 0,\ \mu \neq 0,\ \Delta_1 \neq 0,\ \Delta_2 \neq 0,\ \tau = \Delta_1^2 + \Delta_2^2$$

are integers and $\tau$ is not the square of any rational number. Now

$$\Delta_1(\Delta_1\alpha_1^2 + \tau\Delta_1\alpha_2^2 + 2\tau\alpha_1\alpha_2) = (\alpha_1\Delta_1 + \tau\alpha_2)^2 - \tau\Delta_2^2\alpha_2^2 .$$

Suppose first that $\lambda\Delta_1+\tau\mu=0$. Then, since $\mu\neq0$, $\lambda\neq0$, and $\tau(\mu\Delta_1+\lambda) = \Delta_1(\lambda\Delta_1+\mu\tau)+\lambda\Delta_2^2 = \lambda\Delta_2^2 \neq 0$, we may write $2\Delta_1R$ in the form

$$\begin{aligned} 2\Delta_1R = {} & 2(\alpha_1\Delta_1 + \alpha_2\tau)^2 - 2\tau\Delta_2^2\alpha_2^2 + \lambda\Delta_2^2(\beta_1 + \beta_2)^2 - \lambda\Delta_2^2(\beta_1 - \beta_2)^2 \\ & - \sigma\lambda\Delta_2^2(\beta_3 + \beta_4)^2 + \sigma\lambda\Delta_2^2(\beta_3 - \beta_4)^2, \end{aligned} \tag{54}$$

since $(\beta_1+\beta_2)^2-(\beta_1-\beta_2)^2=4\beta_1\beta_2$. But the right member of (54) satisfies the conditions of Lemma 6 and may be made zero for integers

$$(\alpha_1\Delta_1 + \alpha_2\tau),\quad \alpha_2,\quad (\beta_1 + \beta_2),\quad (\beta_1 - \beta_2),\quad (\beta_3 + \beta_4),\quad (\beta_3 - \beta_4)$$

not all zero and hence for rational numbers $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$ not all zero. It follows that we may make the form (53) vanish for integers $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$ not all zero since it is a homogeneous polynomial in $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$. Next let $\lambda\Delta_1+\tau\mu=\pi\neq0$. Then

$$\begin{aligned} \Delta_1\pi R = {} & \pi(\Delta_1\alpha + \alpha_2\tau)^2 - \pi\tau\Delta_2^2\alpha_2^2 + [\pi\beta_1 + \tau(\mu\Delta_1 + \lambda)\beta_2]^2 \\ & + \tau[\pi^2 - (\mu\Delta_1 + \lambda)^2\tau]\beta_2^2 - \sigma[\pi\beta_3 + \tau(\mu\Delta_1 + \lambda)\beta_4]^2 \\ & - \sigma\tau[\pi^2 - (\mu\Delta_1 + \lambda)^2\tau]\beta_4^2 . \end{aligned} \tag{55}$$

Now $\pi^2-(\mu\Delta_1+\lambda)^2\tau$ is not zero since $\tau$ is not a rational square. Obviously the signs of the coefficients of the squares $(\alpha_1\Delta_1+\alpha_2\tau)^2$ and $\alpha_2^2$ are different and we again have a form satisfying the conditions of Lemma 6. Again we may make (53) vanish for integer values of $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$ not all zero.

LEMMA 7. *We may choose integers* $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$ *not all zero for which the form* $R$ *given by* (53) *vanishes.*

We shall prove

LEMMA 8. *Suppose that* $u$ *is a quantity such that*

$$u^2 = \tau = \Delta_1^2 + \Delta_2^2 \tag{56}$$

*where* $\Delta_1$ *and* $\Delta_2$ *are relatively prime integers and* $\tau$ *is not a rational square. Let* $S$ *be a division algebra which is a generalized quaternion algebra with the basis*

$$1, y, z, yz \tag{57}$$

*and the multiplication table*

$$zy = -yz,\ y^2 = \sigma,\ z^2 = \lambda + \mu u,\ \mu \neq 0,\ \lambda, \mu, \sigma \text{ integers}, \tag{58}$$

*over the field* $R(u)$. *Then* $S$ *contains a quantity* $x$ *satisfying a cyclic quartic equation with integer coefficients.*

For we may choose integers $\alpha_1, \alpha_2, \beta_1, \cdots, \beta_4$ not all zero such that the form (53) vanishes. Then if

$$a_1 = \alpha_1 + \alpha_2 u,\quad b = b_1 + b_2 v,\quad b_1 = \beta_1 + \beta_2 u,\quad b_2 = \beta_3 + \beta_4 u,\quad x = a_1 y + bz$$

we know that $x \neq 0$ and hence $x^2 \neq 0$ since $S$ is a division algebra. We have

$$\begin{aligned} x^2 &= (a_1 y + bz)(a_1 y + bz) = a_1^2 \sigma + bzbz + a_1 b(yz + zy) \\ &= a_1^2 \sigma + (b_1^2 - b_2^2 \sigma)(\lambda + \mu u) = \epsilon_1 + \epsilon_2 u, \end{aligned}$$

where

$$\begin{aligned} \epsilon_1 &= (\alpha_1^2 + \alpha_2^2 \tau)\sigma + \lambda(\beta_1^2 + \beta_2^2 \tau) - \sigma\lambda(\beta_3^2 + \beta_4^2 \tau) + 2\tau\mu(\beta_1\beta_2 - \sigma\beta_3\beta_4), \\ \epsilon_2 &= 2\sigma\alpha_1\alpha_2 + \mu(\beta_1^2 + \beta_2^2 \tau) - \sigma\mu(\beta_3^2 + \beta_4^2 \tau) + 2\lambda(\beta_1\beta_2 - \sigma\beta_3\beta_4). \end{aligned}$$

Consider the linear combination $R = \epsilon_1\Delta_1 + \epsilon_2\tau$. It is obviously the form (53) so that $\epsilon_1\Delta_1 + \epsilon_2\tau = 0$. As we have chosen the $\alpha_i$, $\beta_i$ integers, $\epsilon_1$ is an integer divisible* by $\tau$ and we may write $\epsilon_1 = -\tau\nu$. Then $\epsilon_2 = \nu\Delta_1$ and we have

$$x^2 = \nu(\Delta_1 u - \tau). \tag{59}$$

But $x^2 \neq 0$ so that $\nu \neq 0$. But then

$$\begin{aligned} x^4 &= \nu^2(\Delta_1^2\tau + \tau^2 - 2\tau\Delta_1 u) = \nu^2(\tau\Delta_1^2 + \tau^2) - 2\tau\nu(\nu\tau + x^2) \\ &= \nu^2\tau(\Delta_1^2 - \tau) - 2\tau\nu x^2 = -\nu^2\tau\Delta_2^2 - 2\nu\tau x^2 \end{aligned}$$

and

$$x^4 + 2\nu(\Delta_1^2 + \Delta_2^2)x^2 + \nu^2\Delta_2^2(\Delta_1^2 + \Delta_2^2) = 0. \tag{60}$$

* For when $\Delta_1$ and $\Delta_2$ are relatively prime so are $\tau$ and $\Delta_1$.

The above equation is known to be a canonical form of the cyclic quartic equation with integer coefficients and it may be verified that it has the roots $x$, $\theta(x)$, $\theta^2(x)$, $\theta^3(x)$ with

$$\theta(x) = \frac{x}{\Delta_2}(u + \Delta_1), \quad \theta^4(x) = x.$$

Thus $S$ contains the desired quantity $x$ and our lemma is proved.

We wish next to prove

LEMMA 9. *Let $A$ be a normal division algebra in sixteen units over $R$. Let $A$ contain a quantity $u$ not in $R$ but such that $u^2=\tau$, an integer of $R$. Then the algebra $S$ of all quantities of $A$ which are commutative with $u$ is an algebra of order eight over $R$ and has a basis* (57) *and a multiplication table* (58).

For since every quantity of $S$ is by definition commutative with $u$ not in $R$ and $A$ is a normal division algebra, $A > S$. The order of $S$ is then at most eight. The quantity $-u$ is a root of the minimum equation of $u$ and is a transform $wuw^{-1}$ of $u$ by a quantity $w$ of $A$. Hence $w^2u = w(-u)w = uw^2$ and $w^2$ is in $S$. If $w^2$ is in $R$ the algebra $B=(1, u, w, uw)$ is a generalized quaternion sub-algebra of $A$ which we know to be impossible when $A$ is a division algebra (loc. cit.). Hence the order of $R(w^2)$ is not unity. If the order of $R(w^2)$ is four then the quantity $w$ is in $R(w^2)$ since $A$ has rank four, and $w$ is commutative with $u$, a contradiction. But the order of $R(w^2)$ is a divisor of the order of $R(w)$ and necessarily is two, so that $R(w^2)$ contains a quantity $y$ generating it such that $y^2=\sigma$, a rational integer. The field $R(u, y)$ is a direct product of two quadratic fields and the quantity $i=(u+1)y$ satisfies an irreducible quartic with Galois group $G_4$. Its minimum equation has $-i$ as a root and $-i$ is a transform of $i$ by a quantity $z$ of $A$. In fact $z$ corresponds for this $i$ to the $j_1$ of §2, and the proof there given shows that $z$ is commutative with $u$ and its square is a polynomial $z^2=\lambda+\mu u$ with $\lambda$ and $\mu$ taken to be rational integers without loss of generality. The integer $\mu$ is not zero, for then $A$ would contain a generalized quaternion sub-algebra over $R$. The quantity $y$ corresponds to the $v$ of §2 and $zy=-yz$. We have proved Lemma 9 and by combining it with Lemma 8 and $\tau=\Delta_1^2+\Delta_2^2$ we have shown that $A$ contains a quantity with cyclic quartic minimum equation.

Conversely let $A$ be a cyclic algebra of order sixteen, that is, let $A$ be a linear associative division algebra containing a quantity $x$ with cyclic quartic minimum equation. Since (60) is a canonical form for the cyclic quartic with integer coefficients we may take, by a rational Tschirnhausen transformation, the minimum equation of $x$ in the form (60). By defining $u$ by (59) the minimum equation of $u$ is (56).

THEOREM 3. *Let $A$ be a normal division algebra of order sixteen over $R$. Then $A$ is a cyclic algebra if and only if $A$ contains a quantity $u$ whose minimum equation is*

$$u^2 = \tau = \Delta_1^2 + \Delta_2^2 \tag{56}$$

*where $\Delta_1$ and $\Delta_2$ are rational and $\tau$ is not a rational square.*

For if $\Delta_1$ and $\Delta_2$ were integers but not relatively prime we would define a new quantity $u'$ by $u' = \pi u$ where $\pi$ is the greatest common divisor of $\Delta_1$ and $\Delta_2$ and would have a new $\tau$ with $\Delta_1$ and $\Delta_2$ relatively prime integers. If $\Delta_1$ and $\Delta_2$ were fractions we would reduce the fractions to lowest terms and multiply $u$ by the least common denominator of the fractions and again would have $\Delta_1$ and $\Delta_2$ relatively prime integers by the above process.

We have seen that the constants in the above proof of Lemma 7 were all not merely rational numbers but integers. We may obtain an interesting corollary. Suppose that $A$ contained a quantity $t$ whose minimum equation was

$$\omega^4 + 2\nu_1\tau\omega^2 + \nu_1^2\tau\Delta_4^2 = 0, \quad \tau = \Delta_3^2 + \Delta_4 = \sigma^2\rho, \quad \rho = \Delta_1^2 + \Delta_2^2, \tag{61}$$

with $\tau$ not a product of distinct primes but where $\rho$ is a product of distinct primes. Applying our proof we see that $A$ contains a quantity $x$ with minimum eqation

$$\omega^4 + 2\nu\rho\omega^2 + \nu^2\Delta_2^2\rho = 0, \qquad \rho = \Delta_1^2 + \Delta_2^2, \tag{62}$$

and with $\nu$, $\rho$, $\Delta_1$, $\Delta_2$ all integers. It is easily shown that we may take $\nu$ a product of distinct primes and have

THEOREM 4. *Every cyclic algebra of order sixteen over $R$ contains a quantity $x$ with minimum equation*

$$\phi(\omega) \equiv \omega^4 + 2\nu\rho\omega^2 + \nu^2\Delta_2^2\rho = 0 \tag{63}$$

*where $\nu$ and $\rho$ are each products of distinct primes, $\rho$ is not a rational square, $\rho = \Delta_1^2 + \Delta_2^2$ is any desired representation of $\rho$ as a sum of two integer squares.*

5. **Cyclic algebras.** Every normal division algebra in sixteen units containing a quantity $x$ with minimum equation (63) is known to have a basis

$$x^r y_s \qquad (r,s = 0,1,2,3), \tag{64}$$

and a multiplication table

$$y^r f(x) = f[\theta^r(x)]y^r, \quad y^4 = \gamma \qquad (r = 0,1,2,3) \tag{65}$$

with $\gamma$ an integer of $R$, for every $f(x)$ of $R(x)$. We shall put the cyclic algebra

in our form for algebras of order sixteen (the form of Cecioni). Define $u$ by

$$x^2 = \nu(\Delta_1 u - \rho) \tag{66}$$

so that

$$u^2 = \rho = \Delta_1^2 + \Delta_2^2 \tag{67}$$

where $\rho$ is a product of distinct rational primes and $\rho \neq 0$, $\rho \neq 1$. Obviously the prime factors of $\rho$ have the form 2 or $4n+1$. Also

$$\theta(x) = \frac{-\nu}{\Delta_2}(u + \Delta_1),\ \theta^2(x) = -x,\ \theta^3(x) = -\theta(x),\ \theta^4(x) = x. \tag{68}$$

We have $y^4 = \gamma$, a rational integer having no fourth power as a factor. We may write $\gamma = \gamma_4^3\sigma$ where $\gamma_4$ and $\sigma$ are each products of distinct primes, and $\gamma_4 \neq 0$, $\sigma \neq 0$, $\sigma \neq 1$. It is easily shown that our definition implies that

$$yu = -uy, \quad y^2u = uy^2, \quad xy^2 = -y^2x. \tag{69}$$

Define a quantity $v$ by $v\gamma_4 = y^2$. Then

$$v^2 = \sigma, \quad y^2 = \gamma_4 v, \quad xv = -vx, \quad uv = vu. \tag{70}$$

If $\gamma_4 \equiv 0 \pmod{\rho}$ then $y_1$ defined by $uy_1 = y$ has the property that

$$y_1^2 = -\frac{\gamma_4}{\rho}v, \quad y_1 x = \theta(x)y_1,$$

so that, without loss of generality, we may take $\gamma_4 \not\equiv 0 \pmod{\rho}$. We then have an algebra of the kind constructed in §§2 and 3 and with

$$j_1 = x, \quad j_2 = y, \quad j_3 = xy, \tag{71}$$

so that

$$\begin{aligned} j_1^2 &= -\nu\rho + \nu\Delta_1 u, \quad j_2^3 = \gamma_4 v, \\ j_3^3 &= x\theta(x)y^2 = \nu\Delta_2\gamma_4 uv. \end{aligned} \tag{72}$$

Hence

$$\gamma_1 = -\nu\rho, \ \gamma_2 = \nu\Delta_1, \ \gamma_3 = \gamma_5 = 0, \ \gamma_6 = \nu\Delta_2\gamma_4, \tag{73}$$

and

$$\begin{aligned} &\gamma_1^2 - \gamma_2^2\rho = \nu^2\Delta_2^2\rho, \quad \gamma_3^3 - \gamma_4^3\sigma = -\gamma_4^3\sigma, \\ &\gamma_6^3 - \gamma_6^3\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^3 - \gamma_4^3\sigma). \end{aligned} \tag{74}$$

Conditions (a), (b), (c), (d), (e) of Theorem 2 are satisfied. We let $\rho = \rho_1\pi$, $\sigma = \sigma_1\pi$, where $\pi$ is the greatest common divisor of $\rho$ and $\sigma$. Then we have

THEOREM 5. *Let $\rho_1$, $\sigma_1$, $\pi$ be each products of distinct primes, let these three numbers be relatively prime in pairs, let $\rho_1>1$, $\pi>0$, $\sigma_1\neq 0$, $\sigma_1\neq 1$ and $\rho=\rho_1\pi$ be a product of primes of the form 2 and $4n+1$. Write $\sigma=\sigma_1\pi$ and let $\Delta_1$ and $\Delta_2$ be any particular choice of a pair of integers such that*

$$\rho = \Delta_1^2 + \Delta_2^2 . \tag{75}$$

*Suppose that one of the congruences*

$$\rho \equiv E_1^2 \pmod{\sigma_1}, \quad \sigma \equiv E_2^2 \pmod{\rho_1}, \quad -\sigma_1\rho_1 \equiv E_3^2 \pmod{\pi}$$

*does not hold for any $E_i$. Then the set of algebras given by the basis*

$$x^r y^s \qquad (r,s = 0,1,2,3), \tag{76}$$

*and the multiplication table*

$$x^4 + 2\nu\rho x^2 + \nu^2\Delta_2^2\rho = 0, \quad \theta(x) = \frac{-x}{\Delta_2}(u + \Delta_1), \quad x^2 = \nu(\Delta_1 u - \rho),$$

$$\theta^2(x) = -x, \ \theta^3(x) = -\theta(x), \ y^r a(x) = a[\theta(x)]y^r, \tag{77}$$

$$y^4 = \gamma_4^2\sigma \qquad (a \text{ in } R(x); r = 0,1,2,3),$$

*where $\nu\neq 0$ and $\gamma_4\equiv 0 \pmod{\rho}$ are each products of distinct primes, is a set of associative division algebras over $R$ to one of which every cyclic division algebra of order sixteen over $R$ is equivalent.*

As when $\rho_1$, $\sigma_1$, or $\pi$ is given we know all of its quadratic non-residues we have explicitly determined all cyclic division algebras of order sixteen over $R$.

6. **The algebras of Cecioni.** Cecioni in his Palermo Rendiconti paper constructed normal division algebras based on a non-cyclic abelian equation

$$\psi(\omega) \equiv \omega^4 + p\omega^2 + n^2 = 0.$$

He took $n=1$, $i$ a root of $\psi(\omega)=0$, and had a basis

$$i^r j_s \qquad (r,s = 0,1,2,3; j_0 = 1),$$

where

$$g = j_1^3 = h_1(i^2+1), \ \gamma = j_2^2 = h_2(\theta_2 + i), \ \alpha = \frac{\pm h_1 k(\theta_2 - i)}{g\gamma(\theta_1)},$$

$$g = g(\theta_1), \ j_3^2 = g\gamma\alpha(\theta_1), \ k^2 + c^2 + (\rho+2)h_2^2 = 0$$

so that

$$j_3^2 = \mp h_1 k(\theta_2 - i).$$

Now $(\theta_2-i)^2=\theta_2^2+i^2-2=-(p+2)$, so that if we let $u=\theta_2-i$ then $u^2=-(p+2)$. But $h_2\neq 0$ since otherwise $c=0$ and $j_2^2=0$. Hence $-(p+2)$ is expressible as a sum of two rational squares and we may write

$$u^2 = \tau = \Delta_1^2 + \Delta_2^2 .$$

Our Theorem 3 then shows that $A$ is a cyclic algebra so that *Cecioni's construction did not give any non-cyclic algebras in sixteen units.* We shall take a special case of the algebras, solve the equations necessary, and give the cyclic quantity. Write

$$x = (\theta_2 + i)(u + j_3),$$

and take $k=h_2=1$, $c=4$, $p=-19$, so that

$$-(p+2) = 17 = 1 + 4^2, \ \Delta_1 = 1, \ \Delta_2 = 2.$$

Then, since $j_3(\theta_2+i) = -(\theta_2+i)j_3$, $j_3u=uj_3$, $j_3^2=u$,

$$x^2 = (\theta_2 + i)^2u^2 + [(\theta_2 + i)j_3]^2 = 21\cdot 17 - 21u,$$

and

$$x^2 = -21(u - 17).$$

It follows that

$$x^4 + 2(-21)\cdot 17x^2 + (21)^2\cdot(2)^2\cdot 17 = 0,$$

a cyclic quartic with

$$x^2 = -21(u - 17), \quad \theta(x) = (x/2)(u + 1).$$

The algebras of Cecioni have been shown to be cyclic algebras and no non-cyclic algebras are known to exist. Cecioni's choice of $\rho$, $\sigma$ made his algebras cyclic immediately no matter what the $\gamma_1, \cdots, \gamma_6$ are. The conditions that $A$ be an associative algebra may be satisfied for $\rho$, $\sigma$ not so obviously making $A$ a cyclic algebra. For example we may take $\rho=3$, $\sigma=-1$ so that neither $\rho$, $\sigma$ nor $\sigma\rho$ is a sum of two rational squares. Then if $\gamma_1=3$, $\gamma_2=\gamma_3=\gamma_4=1$, $\gamma_5=0$, $\gamma_6=2$, we have

$$\gamma_1^2 - \gamma_2^2\rho = 6, \ \gamma_3^2 - \gamma_4^2\sigma = 1 + 1 = 2, \ \gamma_5^2 - \gamma_6^2\sigma\rho = 12,$$

and $A$ is associative. But $A$ is a division algebra since

$$6 \neq \mu_1^2 + \mu_2^2$$

for any rational $\mu_1$ and $\mu_2$. But, as we shall show later, even this algebra is a cyclic algebra. We are led to an investigation of the diophantine conditions on $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ that $A$ be a non-cyclic algebra.

7. **A set of diophantine necessary and sufficient conditions that $A$ be a non-cyclic algebra.** We shall consider a normal division algebra in sixteen units over $R$, and shall assume that neither $\rho$, $\sigma$, nor $\sigma\rho$ is expressible rationally in the form $\Delta_1^2+\Delta_2^2$. Then it is true that no polynomial $s$ of $R(i)$ which is not in $R$, that is, which is not merely a rational number, has the property that

$$s^2 = \Delta_1^2 + \Delta_2^2 \qquad (\Delta_1 \text{ and } \Delta_2 \text{ in } R). \tag{78}$$

For let $s$ be a quantity of $R(i)$ which is not in $R$ but whose square is in $R$. Then $s=s_1+s_2v$ and $s^2=s_1^2+s_2^2\sigma+2s_1s_2v$. If $s^2$ is in $R$ then $s_1s_2=0$ and $s_1=0$ or $s_2=0$. If $s_2=0$ then $s_1=\lambda_1+\lambda_2u$ and $s^2=s_1^2=\lambda_1^2+\lambda_2^2\rho+2\lambda_1\lambda_2u$. But $s$ is not in $R$ so that when $\lambda_1\lambda_2=0$ then $\lambda_1=0$ and $s=\lambda_2u$. If $s_2\neq0$, $s_1=0$ and $s=(\lambda_3+\lambda_4u)v$, $s^2=(\lambda_3^2+\lambda_4^2\rho+2\lambda_3\lambda_4u)\sigma$. Hence $\lambda_3\lambda_4=0$. If $\lambda_4=0$ then $s=\lambda_3v$ while if $\lambda_3=0$ then $s=\lambda_4uv$. We have proved

LEMMA 10. *The only quantities of $R(i)$ which are not in $R$ but are such that their squares are in $R$ are those quantities of the form*

$$s = \lambda_2u,\ \lambda_3v,\ \text{or}\ \lambda_4uv, \tag{79}$$

*so that*

$$s^2 = \lambda_2^2\rho,\ \lambda_3^2\sigma,\ \text{or}\ \lambda_4^2\sigma\rho. \tag{80}$$

But by hypothesis neither $\rho$, $\sigma$ nor $\sigma\rho$ is a sum of two rational squares and we have the desired result. As a corollary of Lemma 10 we have

LEMMA 11. *The only quantities of grade two in $R(i)$ are those quantities of the form*

$$t = \lambda_1 + \lambda_2u,\ \lambda_1 + \lambda_3v,\ \text{or}\ \lambda_1 + \lambda_4uv. \tag{81}$$

We wish now to find necessary and sufficient conditions that the algebra $\Sigma_1$ of all quantities of the form

$$a + bj_1 \qquad (a,b \text{ in } R(i)) \tag{82}$$

contain a quantity $s$ such that $s^2=\Delta_1^2+\Delta_2^2$. If this be true then $s=a+bj_1$, $b\neq0$,

$$s^2 = a^2 + (bj_1)^2 + b[a + a(-v)]j_1, \tag{83}$$

so that, by the linear independence of the basal units of $A$, we have

$$b[a + a(-v)]j_1 = 0.$$

But $b\neq0$ and thus $a=a_2+a_1v=-(a_2-a_1v)=a_1v$ with $a_1$ in $R(u)$. Also if

(84) $$b = \beta_1 + \beta_2 u + (\beta_3 + \beta_4 u)v \qquad (\beta_1, \cdots, \beta_4 \text{ in } R)$$

then

(85) $$(bj_1)^2 = f_1 + f_2 u,$$

where

(86) $$f_1 = \gamma_1[(\beta_1^2 + \beta_2^2\rho) - \sigma(\beta_3^2 + \beta_4^2\rho)] + 2\gamma_2\rho(\beta_1\beta_2 - \sigma\beta_3\beta_4),$$

(87) $$f_2 = \gamma_2[(\beta_1^2 + \beta_2^2\rho) - \sigma(\beta_3^2 + \beta_4^2\rho)] + 2\gamma_1(\beta_1\beta_2 - \sigma\beta_3\beta_4).$$

Then if $a_1 = \alpha_1 + \alpha_2 u$ we have

$$s^2 = (\alpha_1^2 + \alpha_2^2\rho + 2\alpha_1\alpha_2 u)\sigma + f_1 + f_2 u$$

and $s^2 = \Delta_1^2 + \Delta_2^2$ if and only if

(88) $$(\alpha_1^2 + \alpha_2^2\rho)\sigma + f_1 = \Delta_1^2 + \Delta_2^2, \ 2\alpha_1\alpha_2\sigma + f_2 = 0.$$

We may eliminate $\alpha_2$ between the two equations of (81) and, using the fact that $R(i)$ contains no quantities $s$ with the desired property, we know that $\beta_1, \cdots, \beta_4$ are not all zero. This gives Lemma 12.

LEMMA 12. *Let $A$ be a normal division algebra of order sixteen. Then its sub-algebra $\Sigma_1$ whose quantities are of the form $a + bj_1$ contains a quantity $s$, not in $R$ but such that $s^2 = \Delta_1^2 + \Delta_2^2$ with $\Delta_1$ and $\Delta_2$ rational, if and only if there exist rational integers $\alpha_1, \beta_1, \cdots, \beta_4, \Delta_1, \Delta_2$ such that $\beta_1, \cdots, \beta_4$ are not all zero and*

(89) $$f_2^2\rho + 4\alpha_1^4\sigma^2 + 4\alpha_1^2\sigma(f_1 - \Delta_1^2 - \Delta_2^2) = 0.$$

For if $\alpha_1 = 0$ and (82) is satisfied then $f_2 = 0$ contrary to the hypothesis that $A$ is a division algebra. Hence $\alpha_1 \neq 0$ and (89) is equivalent to (88). Since (89) is a homogeneous polynomial in the variables, its solution in rational numbers is equivalent to its solution in integers.

By symmetry we have

LEMMA 13. *Sub-algebra $\Sigma_2$ of division algebra $A$, containing quantities $a + bj_2$ with $a$ and $b$ in $R(i)$, contains a quantity $s$ not in $R$ and yet such that $s^2 = \Delta_1^2 + \Delta_2^2$ if and only if*

(90) $$f_4^2\sigma + 4\alpha_1^4\rho^2 + 4\alpha_1^2\rho(f_3 - \Delta_1^2 - \Delta_2^2) = 0$$

*for integers $\alpha_1, \Delta_1, \Delta_2$, and $\beta_1, \cdots, \beta_4$ not all zero, where $f_3$ is given by* (36) *and $f_4$ by* (37).

LEMMA 14. *Sub-algebra $\Sigma_3$ of division algebra $A$, containing quantities $a+dj_3$ with $a$ and $d$ in $R(i)$, contains a quantity $s$ not in $R$ and yet such that $s^2=\Delta_1^2+\Delta_2^2$ if and only if*

$$f_6^2\sigma\rho + 4\alpha_1^4\rho^2 + 4\alpha_1^2\rho(f_5 - \Delta_1^2 - \Delta_2^2) = 0 \tag{91}$$

*for integers $\alpha_1$, $\Delta_1$, $\Delta_2$ and $\delta_1, \cdots, \delta_4$ not all zero, where $f_5$ is given by* (38) *and $f_6$ by* (39).

We shall consider, at this point, the special example given at the end of §5. In that example we had $\sigma=-1$, $\rho=\gamma_1=3$, $\gamma_2=1$. In that case take $\beta_1=\beta_3=\beta_4=0$, $\beta_2=\alpha_1=2$ so that $f_2=3\beta_2^2$,

$$f_2^2\rho = 27\beta_2^4 = 27\alpha_1^4,\ 4\alpha_1^2\sigma f_1 = -36\alpha_1^4,$$

and

$$f_2^2\rho + 4\alpha_1^4\sigma^2 + 4\alpha_1^2\sigma f_1 = -5\alpha_1^4.$$

But then if $\Delta_1=1$, $\Delta_2=2$,

$$4\alpha_1^2\sigma(-\Delta_1^2-\Delta_2^2) = 4, \quad 5\alpha_1^2 = 5\alpha_1^4,$$

and the form (89) vanishes so that algebra $\Sigma_1$ contains a quantity

$$s = (2+3u)v + 6uj_1, \quad s^2 = 5,$$

since $2\alpha_1\alpha_2\sigma=-f_2$, $\alpha_2=3$. Then algebra $A$ is cyclic.

Let us now assume that neither $\Sigma_1$, $\Sigma_2$ nor $\Sigma_3$ contains a quantity $s$ such that $s^2=\Delta_1^2+\Delta_2^2$. We wish to investigate the conditions that a quantity

$$s = T_1 + Tj_2 \qquad (T_1, T \neq 0 \text{ in } \Sigma_1) \tag{92}$$

be such that $s^2=\Delta_1^2+\Delta_2^2$ with $\Delta_1$ and $\Delta_2$ in $R$. If $s^2$ is in $R$, then since

$$s^2 = T_1^2 + (Tj_2)^2 + T_1(Tj_2) + (Tj_2)T_1$$

and since

$$(Tj_2)^2 = TT'g_2, \quad (Tj_2)T_1 = TT_1'j_2,$$

with $T_1'$ and $T'$ in $\Sigma$, we have

$$T_1(Tj_2) + (Tj_2)T_1 = 0, \quad T_1^2 + W = \Delta_1^2 + \Delta_2^2, \tag{93}$$

where

$$W = (Tj_2)^2 = TT'g_2. \tag{94}$$

Since $W$ is in $\Sigma_1$ and $Tj_2$ is not in $\Sigma_1$ we know that $Tj_2$ is not a polynomial in $W$ and the order of $R(W)$ is less than that of $R(Tj_2)$. But $W$ is not in $R$ since then $(Tj_2)^2$ would be in $R$ and $A$ would contain a generalized quaternion

sub-algebra $(1, u, Tj_2, uTj_2)$. Hence $W$ has grade two and $R(W)$ contains $T_1^2 = -W + \Delta_1^2 + \Delta_2^2$. The field $R(T_1)$ contains $W$. Since $T_1$ is in $\Sigma_1$ an algebra over $R(u)$, $T_1 u = uT_1$. The field $R(u, W)$ has order four since $u$ is not in $R(W) > R(Tj_2)$. Hence $R(T) = R(u, T_1^2) = R(u, W)$, for $T$ must be in any field of order four with whose quantities it is commutative.* Then

$$T_1 = P_1(W) + P(W)u,$$

and since $(Tj_2)T_1 = -T_1(Tj_2)$ we have $P_1(W) = 0$ and $s = P(W)u + Tj_2$ with $P(W) = \lambda_1 + \lambda_2 W$, $\lambda_1$ and $\lambda_2$ in $R$, and $T$ in $\Sigma_1$. Since $W$ has grade two we may write

$$W^2 = 2\phi_1 W + \phi_2, \quad V = W - \phi_1, \quad V^2 = \phi_2 + \phi_1^2 \equiv \psi,$$

and if $P(W) = \alpha_1 + \alpha_2 V$, $W = V + \phi_1$, then

$$s^2 = (\alpha_1^2 + \alpha_2^2 \psi + 2\alpha_1\alpha_2 V)\rho + V + \phi_1,$$

which equals $\Delta_1^2 + \Delta_2^2$ if and only if

$$(\alpha_1^2 + \alpha_2^2\psi)\rho + \phi_1 = \Delta_1^2 + \Delta_2^2, \quad 2\alpha_1\alpha_2\rho = -1. \tag{95}$$

It now becomes necessary to compute $\phi_1$ and $\phi_2$. We may write

$$Tj_2 = (b + dj_1)j_2 = bj_2 + dj_3 \qquad (b, d \text{ in } R(i)).$$

If $d = 0$ then $bj_2$ is in $\Sigma_2$ so that $s$ is in $\Sigma_2$ contrary to our hypothesis that $\Sigma_2$ contains no quantities $s$ such that $s$ is not in $R$ and yet $s^2 = \Delta_1^2 + \Delta_2^2$. Similarly $b \neq 0$ and we shall now write

$$(Tj_2)^2 = W = (bj_2)^2 + (dj_3)^2 + bd(-u)j_2j_3 + db(-u, -v)j_3j_2.$$

We have taken $A$ so that $j_2j_3 = \alpha g_2(-v)j_1$, $j_3j_2 = g_2(-v)g_1$ and

$$W = (bj_2)^2 + (dj_3)^2 + [b\alpha(-u) + db(-u, -v)]g_2(-v)j_1.$$

Write

$$(bj_2)^2 = f_3 + f_4 v, \quad (dj_3)^2 = f_5 + f_6 uv,$$

where, as we know, $f_3$, $f_4$, $f_5$ and $f_6$ are given respectively by (36), (37), (38), (39). Then

$$W = q_1 + q_2 j_1, \tag{96}$$

with

$$q_1 = (f_3 + f_5) + f_4 v + f_6 uv, \quad q_2 = [bd(-u)\alpha + db(-u, -v)]g_2(-v) \tag{97}$$

and

* It is known that the only quantities of a normal division algebra of order $n^2$ over $F$ which are commutative with a quantity $x$ of grade $n$ in the algebra, are polynomials in $x$ with coefficients in $F$.

$$g_1 g_2(-v)\alpha = g_3(-v). \tag{98}$$

Then

$$W^2 = q_1^2 + (q_2 j_1)^2 + q_2[q_1 + q_1(-v)]j_1.$$

We shall prove

LEMMA 15. *The quantity $q_2$ is zero if and only if $b=0$ or $d=0$.*

For let $b=0$ or $d=0$. By its form $q_2$ vanishes. Let next $b\neq 0$ and $d\neq 0$. If $q_2=0$ then

$$W = f_3 + f_5 + (f_4 + f_6 u)v \quad \text{in} \quad R(i).$$

But $W$ has grade two with respect to $R$ as we have shown. Also the only polynomials in $i$ of grade two were proved in Lemma 11 to be quantities of the form $\lambda_1+\lambda_2 u$, $\lambda_1+\lambda_3 v$, $\lambda_1+\lambda_4 uv$. If $W$ has any of these three forms then $f_4=0$ or $f_6=0$. But when $A$ is a division algebra $f_4=0$ if and only if $b=0$ and $f_6$ was shown to be zero if and only if $d=0$. This furnishes a contradiction and Lemma 15 is true.

We have assumed that $b\neq 0$ and $d\neq 0$. Then $q_2\neq 0$, and since

$$q_1^2 + (q_2 j_1)^2 + q_2[q_1 + q_1(-v)]j_1 = 2\phi_1(q_1 + q_2 j_1) + \phi_2,$$

we have

$$2\phi_1 = q_1 + q_1(-v) = 2(f_3 + f_5), \quad \phi_1 = f_3 + f_5, \tag{99}$$

while

$$\phi_2 = q_1^2 + (q_2 j_1)^2 \phi_1 q_1 = (q_2 j_1)^2 - q_1 q_1(-v).$$

We may easily compute $q_1 q_1(-v)$ and have

$$q_1 q_1(-v) = (f_3 + f_5)^2 - \sigma(f_4^2 + f_6^2 \rho + 2 f_4 f_6 u). \tag{100}$$

To compute $(q_2 j_1)^2$ replace $\alpha$ by its value and we have

$$g_1 q_2 = bd(-u)g_3(-v) + db(-u, -v)g_1 g_2(-v).$$

Also

$$g_1(q_2 j_1)^2 = g_1 q_2 q_2(-v)g_1.$$

But

$$g_1 q_2(-v) = b(-v)b(-u, -v)g_3 + d(-v)b(-u)g_1 g_2$$

and

$$\begin{aligned} g_1(q_2 j_1)^2 = {} & [bb(-v)d(-u)d(-u, -v)g_3 g_3(-v) \\ & + dd(-v)b(-u)b(-u, -v)g_1^2 g_2 g_2(-v)] \\ & + g_1[bb(-u)d(-u)d(-v)g_2 g_3(-v) \\ & + b(-v)b(-u, -v)dd(-u, -v)g_2(-v)g_3]. \end{aligned} \tag{101}$$

But $g_3g_3(-v)=g_1g_1(-u)g_2g_2(-v)$ so that the right member of (101) has a factor $g_1$. First

$$[bb(-u)g_2][d(-u)d(-v)g_3(-v)] = (f_3+f_4v)(f_5-f_6uv),$$

so that the second term in brackets of the right member of (101) is

$$2(f_3f_5-f_4f_6\sigma u)g_1. \tag{102}$$

The first term has a factor $g_1g_2g_2(-v)$, and apart from this factor is

$$g_1b(-u)b(-u,-v)dd(-v)+bb(-v)d(-u)d(-u,-v)g_1(-u). \tag{103}$$

Also since

$$b=\beta_1+\beta_2v+(\beta_3+\beta_4v)u=\beta_1+\beta_3u+(\beta_2+\beta_4u)v,$$
$$d=\delta_1+\delta_2uv+(\delta_3+\delta_4uv)u=\delta_1+\delta_3u+(\delta_4\rho+\delta_2u)v,$$

we have

$$bb(-v)=\beta_1^2+\beta_3^2\rho-(\beta_2^2+\beta_4^2\rho)\sigma+2(\beta_1\beta_3-\sigma\beta_2\beta_4)u,$$
$$dd(-v)=\delta_1^2+\delta_3^2\rho-(\delta_2^2+\delta_4^2\rho)\sigma\rho+2(\delta_1\delta_3-\sigma\rho\delta_2\delta_4)u.$$

Let

$$\begin{aligned} h_3&=\beta_1^2+\beta_3^2\rho-\sigma(\beta_2^2+\beta_4^2\rho),\\ h_4&=-2(\beta_1\beta_3-\sigma\beta_2\beta_4),\\ h_5&=\delta_1^2+\delta_3^2\rho-\sigma\rho(\delta_2^2+\delta_4^2\rho),\\ h_6&=2(\delta_1\delta_3-\sigma\rho\delta_2\delta_4). \end{aligned} \tag{104}$$

Then $b(-u)b(-u,-v)=h_3+h_4u$ and

$$g_1dd(-v)b(-u)b(-u,-v)=(h_3+h_4u)(h_5+h_6u)(\gamma_1+\gamma_2u).$$

But (103) is a polynomial in $u$ added to the same polynomial in $-u$ and its value is

$$H=2[\gamma_1(h_3h_5+h_4h_6\rho)+\gamma_2\rho(h_3h_6+h_4h_5)], \tag{105}$$

so that

$$\phi_2=\sigma f_4^2+\sigma\rho f_6^2-f_3^2-f_5^2+(\gamma_3^2-\gamma_4^2\sigma)H.$$

But

$$\psi=\phi_2+\phi_1^2=\phi_2+f_3^2+f_5^2+2f_3f_5$$

so that

$$\psi=\sigma f_4^2+\sigma\rho f_6^2+2f_3f_5+(\gamma_3^2-\gamma_4^2\sigma)H. \tag{106}$$

The form (106) reduces to $\sigma f_4^2$ when $d=0$ and to $\sigma\rho f_6^2$ when $b=0$. Also $s^2=\Delta_1^2+\Delta_2^2$ if and only if

$$\psi + 4\alpha_1^4\rho^2 + 4\alpha_1^2\rho(\phi_1 - \Delta_1^2 - \Delta_2^2) = 0, \tag{107}$$

for if $\alpha_1=0$ then $V^2=\psi=0$ which is not true since, as we have shown, $W$ is not in $R$. Hence (107) is equivalent to (95). But it reduces to (90) when $d=0$ since then $f_5=0$, $\phi_1=f_3$. Similarly (107) reduces to (91) for $b=0$. We then have (90) and (91) included in (107) and may omit them. Suppose that (107) were satisfied for values of the variables not all zero. If all of the $\beta_i$ and $\delta_i$ were zero then we could, by taking the remaining variables not all zero, have a solution. But the resulting quantity $s$ would be zero. We must therefore seek solutions with not all of the $\beta_i$ and $\delta_i$ vanishing. The solution of (107) in rational numbers is equivalent to its solution in integers and we have

THEOREM 6. *A normal division algebra $A$ in sixteen units over $R$ with integer parameters $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ satisfying the conditions of Theorem 2, and such that neither $\rho$, $\sigma$ nor $\sigma\rho$ is a sum of two integer squares, is a non-cyclic algebra if and only if*

$$\text{(a)} \qquad f_2^2\rho + 4\alpha_1^4\sigma^2 + 4\alpha_1^2\sigma(f_1 - \Delta_1^2 - \Delta_2^2) \neq 0,$$

$$\text{(b)} \qquad \psi + 4\alpha_1^4\rho^2 + 4\alpha_1^2\rho(\phi_1 - \Delta_1^2 - \Delta_2^2) \neq 0,$$

*for any integers $\alpha_1$, $\Delta_1$, $\Delta_2$, $\beta_1, \cdots, \beta_4$, $\delta_1, \cdots, \delta_4$ such that $\beta_1, \cdots, \beta_4$, $\delta_1, \cdots, \delta_4$ are not all zero and $f_1$ is given by* (86), *$f_2$ by* (87), *$\phi_1$ by* (99), (36), *and* (38), *and $\psi$ by* (106), (105), (104), (37), (39).

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# ON COMPOSITION OF SINGULARITIES*

BY

W. J. TRJITZINSKY

1. **Introduction.** Our purpose will be to establish several theorems on composition of singularities analogous to Hadamard's multiplication theorem.† In each of these we consider essentially an integral of the form

$$F(z) = \frac{1}{2\pi i}\int_C f(x)\cdot g(q(z,x))\cdot dx/x. \tag{1}$$

(In the proof of Hadamard's theorem (1) is used, where $q(z, x)=z/x$.) One of our theorems (§2) is derived with $q(z, x)=\log z/\log x$; another (§3), which is derived with $q(z, x)=(ax+bz+c)/(a_1x+b_1z+c_1)$, is a generalization of Hadamard's theorem. Further, the theorem of §3 is extended to functions of two complex variables (§4) in a way analogous to the extension by Haslam-Jones‡ of the one-variable Hadamard's theorem to functions of two variables. Finally, a theorem is derived (§5) making use of (1), with $q(z, x)$ a more general function than the one employed in §3.

2. **The case where $q(z, x)=\log z/\log x$.** We shall denote the closed region of the complex plane, consisting of all those points $x$ for which $\alpha \leqq \arg x \leqq 2\pi-\alpha$ $(0<\alpha<\pi)$, by $W(\alpha)$. Let $f_1(x_1)$ be a uniform analytic function whose singularities $\alpha_i'$, $0<|\alpha_1'|\leqq|\alpha_2'|\leqq\cdots$, form an isolated set, and let $g_1(x_1)$ be another uniform analytic function with a single singularity $\beta'$, $|\beta'|>0$. By simple transformations $f_1(x_1)$ can be changed to $f(x)$, a uniform analytic function, whose singularities $\alpha_i$, $1<|\alpha_1|\leqq|\alpha_2|\leqq\cdots$, form an isolated set such that none of them belong to $W(\phi)$ $(0<\phi<\pi)$. Since the $\alpha_i'$ are isolated, such a region $W(\phi)$ can be found. Let $\phi_1$ and $\beta$ be defined by the following relations:

$$\pi/\beta = \phi_1, \quad \phi < \phi_1 < \pi. \tag{1}$$

Then transform $g_1(x_1)$ so that the new function $g(x)$ is uniform, analytic and has only one singularity $\beta$. By (1), $\beta>1$ and is real. Without any loss of generality we may consider $f(x)$ and $g(x)$ in place of $f_1(x_1)$ and $g_1(x_1)$.

Consider in the $x$-plane a region $G$ consisting of the portion of the plane

* Presented to the Society, March 29, 1929; received by the editors April 26, 1929.

† Hadamard, Acta Mathematica, vol. 22 (1898), p. 55. Faber, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 16 (1907), p. 285.

‡ Haslam-Jones, Proceedings of the London Mathematical Society, vol. 27 (1927), p. 223.

bounded by a circle $C_R$, which has radius $R$ and center $x=0$, and which does not pass through any of the $\alpha_i$, the points of $W(\phi)$ and the neighborhood of the origin being excluded, the latter by means of a small circle $C_0$ with $x=0$ for the center and radius $r_0$. For $R$ sufficiently large a number of the points $\alpha_i$ will lie within the circle $C_R$. Surround each of these by small circles $C_i$, with these points for centers and with radii $r_i$, respectively. Join each $C_i$ with $C_R$ by means of a Jordan curve $l_i$.* The interiors of the $C_i$ do not belong to $G$. Choose the $r_i$ sufficiently small so that $C_i$ has no points in common with $C_k$, $C_R$, $C_0$ or the boundary of $W(\phi)$. Also the $l_i$ can be taken so that $l_i$ has no points in common with $l_k$, $C_k$, $C_0$ or the boundary of $W(\phi)$. The region $G$, thus defined, is evidently a simply connected region. Its boundary $C$ consists of some of the circumferences $C_i$, the curves $l_i$, of those portions of the circumferences $C_R$, $C_0$ which are left when the points of $W(\phi)$ are excluded from them, and of the part of the boundary of $W(\phi)$ interior to $C_R$ and exterior to $C_0$ (this part consists of portions of straight lines through the origin).

Consider the transformation

$$z = x^\beta \tag{2}$$

and apply it to the contour $C$. A circumference $C_i$ is given by

$$x = \alpha_i + r_i e^{i\theta} \qquad (0 \leqq \theta \leqq 2\pi). \tag{3}$$

Its transformation in the $z$-plane will be given by a contour $C_{i\beta}$,

$$\begin{aligned} z &= (\alpha_i + r_i e^{i\theta})^\beta = \alpha_i^\beta (1 + r_i e^{i\theta}/\alpha_i)^\beta, \\ z - \alpha_i^\beta &= r_i \beta \alpha_i^{\beta-1} e^{i\theta}[1 + r_i p(\theta)] \qquad (0 \leqq \theta < 2\pi), \end{aligned} \tag{4}$$

where $p(\theta)$ is of period $2\pi$ in $\theta$ and is bounded so that with $r_i$ sufficiently small $0<1-\epsilon<|1+r_i p(\theta)|<1+\epsilon$. Thus it follows from (4) that $C_{i\beta}$ is a simple closed curve containing $z=\alpha_i^\beta$ and with radii

$$(1-\epsilon) r_i \beta \,|\, \alpha_i^{\beta-1} \,| \quad \text{and} \quad (1+\epsilon) r_i \beta \,|\, \alpha_i^{\beta-1} \,|,$$

respectively. When $r_i$ is made to approach zero, the contour $C_{i\beta}$ will reduce to the point $z=\alpha_i^\beta$.

The portion of the circumference $C_0$ belonging to $C$ consists of points $x$ given by

$$x = r_0 e^{i\theta}, \quad -\phi < \theta < \phi.$$

Its transformation $C_{0\beta}$,

$$z = r_0^\beta e^{i\beta\theta} = r_0^\beta e^{i\theta_1}, \quad -\beta\phi < \theta_1 < \beta\phi, \tag{5}$$

* Compare the procedure in this section with the proof of Hadamard's theorem, as given by Mandelbrojt; S. Mandelbrojt, The Rice Institute Pamphlet, vol. 14 (1927), No. 4, pp. 242–245.

is an arc of a circle in the $z$-plane. By (1), $\beta\phi<\pi$. Similarly, the transformation of the part of $C_R$ belonging to $C$ is an arc of a circle $C_{R\beta}$. This arc consists of the points

$$z = R^{\beta}e^{i\theta_1}, \quad -\beta\phi < \theta_1 < \beta\phi < \pi. \tag{6}$$

The part of the boundary of $W(\phi)$ belonging to $C$,

$$x = |x|e^{\pm i\phi}, \quad r_0 \leqq |x| \leqq R, \tag{7}$$

is transformed into

$$z = |x|^{\beta}e^{\pm i\beta\phi}, \quad r_0^{\beta} \leqq |x^{\beta}| \leqq R^{\beta}, \tag{8}$$

or

$$\arg z = \pm\beta\phi, \quad r_0^{\beta} < |z| < R^{\beta}. \tag{9}$$

The Jordan curves $l_i$ are transformed into Jordan curves $l_{i\beta}$, where $l_{i\beta}$ joins $C_{i\beta}$ with $C_{R\beta}$. Denote the transformation of $C$ by $C_\beta$. With the $r_i$ sufficiently small, $C_\beta$ is the boundary of a simply connected region $G_\beta$. This boundary consists of points $z$, satisfying (5), (6), (9), and of the curves $C_{i\beta}$ and $l_{i\beta}$.

Throughout this section $\log u$ will denote the branch which reduces to zero for $u=1$. The $r_i$ will be taken small enough and the $l_i$ be so chosen that $x=1$ is not on $C$. Consider the following integral:

$$F(z) = \frac{1}{2\pi i}\int_C f(x)\cdot g\left(\frac{\log z}{\log x}\right)\frac{dx}{(x-1)^p}, \tag{10}$$

where $p$ is a positive integer or zero. Since $x$ is on $C$, for $z$ interior to $C_\beta$ $\log z/\log x \neq \beta$. If it were otherwise it would follow that

$$z = x^{\beta};$$

and since $x$ is a point of $C$, $z$ would be on $C_\beta$. For $z$ within $C_\beta$ and $x$ on $C$, $f(x)$ is regular in $x$ and $g(\log z/\log x)$ is continuous in $z$ and $x$. Consequently $F(x)$ is continuous. Since $F^{(1)}(z)$ exists on account of the existence and continuity of $\partial g(\log z/\log x)/\partial z$, it follows that $F(z)$ is analytic within $C_\beta$. By letting $R$ approach infinity in such a way that the $\alpha_i$ are never on $C$, and by diminishing the $r_i$ indefinitely, the circular portion of the contour $C_\beta$, which belongs to $C_{R\beta}$, will be made to recede to infinity and the simple closed curves $C_{i\beta}$ will be made to close down on the points $\alpha_i^{\beta}$, respectively. The $l_i$ and consequently the $l_{i\beta}$ are arbitrary. Let $L(\phi\beta)$ denote the set complementary to $W(\phi\beta)$ and $C_{0\beta}$. In $L(\phi\beta)$, $F(z)$ is uniform and analytic with a possible exception of points of the form $\alpha_i^{\beta}$.

We shall show that when $z$ is restricted to a certain subregion $L$ of $C_\beta$,

the integrand of (10) is a uniform and analytic function of $x$ in a region $M$ consisting of the circumference of a small circle $C'$ of radius $r'$

$$x = 1 + r'e^{i\theta} \qquad (0 \leqq \theta < 2\pi)$$

and the open region $O$ bounded by $C'$ and $C$. For that purpose it is sufficient to prove the following lemma:

LEMMA I. *Let $a, b, c$ be positive numbers defined in succession by the following conditions*:

(i) *For $x$ in $M$, $\beta|\log x| > a$*; (ii) *$b$ is sufficiently small, so that $a^2 - b^2 > 0$*; (iii) $c^2 < a^2 - b^2$.

*Let a region $L$ consist of the points*

(iv) $1 \leqq |z| < e^c, \qquad |\arg z| \leqq b,$

*which lie within $C_\beta$ (such a region exists since there is a neighborhood of $z=1$ interior to $C_\beta$).*

*Then*

$$|\log z/\log x| < \beta \tag{11}$$

*for all $z$ in $L$ and $x$ in $M$.*

A number $a$ satisfying (i) exists since there is a neighborhood of $x=1$ which is not a part of $M$. From (iv) it follows that

$$0 \leqq \log|z| < c.$$

Hence

$$|\log z| < (c^2 + \phi^2)^{1/2}, \quad \phi = \arg z.$$

By (iv) this gives

$$|\log z| < (c^2 + b^2)^{1/2}$$

for all $z$ in $L$.

Thus we have

$$|\log z| < (c^2 + b^2)^{1/2}, \quad a < \beta \cdot |\log x|.$$

By (iii), $(c^2+b^2)^{1/2} < a$. This gives

$$|\log z| < \beta \cdot |\log x|$$

or (11). This proves the lemma.

For $z$ fixed within the region $L$ defined by Lemma I, the integrand (10) is a uniform and analytic function of $x$ in $M$. The contour $C$ can thus be deformed into $C'$ without encountering any singular points. We have therefore

$$F(z) = \frac{1}{2\pi i}\int_C f(x)\cdot g\left(\frac{\log z}{\log x}\right)\cdot\frac{dx}{(x-1)^p} \tag{12}$$

$$= \frac{1}{2\pi i}\int_{C'} f(x)\cdot g\left(\frac{\log z}{\log x}\right)\cdot\frac{dx}{(x-1)^p}.$$

Let

$$f(x) = \sum_{n=0}^{\infty} a_n\cdot(x-1)^n, \tag{13}$$

where the power series converges within a circle

$$x = (|\alpha_1| - 1)e^{i\theta} \qquad (0 \leqq \theta < 2\pi),$$

and let

$$g(x) = \sum_{n=0}^{\infty} b_n\cdot x^n \qquad (|x| < \beta). \tag{14}$$

For purposes of computation (12) can be written in the form

$$F(z) = \frac{1}{2\pi i}\int_{C'} \sum a_n(x-1)^n\cdot\sum b_n\left(\frac{\log z}{\log x}\right)^n\cdot\frac{dx}{(x-1)^p}. \tag{15}$$

We have thus proved the following theorem:

THEOREM I. *Given the series*

$$\sum a_n(x-1)^n, \quad \sum b_n x^n$$

*representing the uniform and analytic functions $f(x)$, $g(x)$, respectively. Let the singularities $\alpha_i$ of $f(x)$ form an isolated set such that $1<|\alpha_i|$ and that none of the $\alpha_i$ are in $W(\phi)$, $\phi<\pi$. Let the only singularity of $g(x)$ be $\beta$, $\beta=\pi/\phi_1$, $\phi<\phi_1<\pi$. With a possible exception of points of the form $\alpha_i^{\beta}$, the function*

$$F(z) = \frac{1}{2\pi i}\int_{C'} f(x)\cdot g\left(\frac{\log z}{\log x}\right)\cdot\frac{dx}{(x-1)^p} \tag{16}$$

*($p$ a positive integer) is uniform and analytic in the region $L(\phi\beta)$. This region consists of the set complementary to $W(\phi\beta)$ and to a small circle around the origin. ($C'$ is a small circle around $x=1$.)*

**Note.** Using (15) it can be shown that

$$F(z) = \sum_{n=0}^{\infty} c_n'(\log z)^n = \sum c_n(z-1)^n, \tag{17}$$

where the power series converges within a circle having $z=1$ for center and a radius $\rho$ where $\rho<1$ and $\rho<|\alpha_i-1|$ for all $i$. The $c_n$ depend on the $a_i$ and the $b_m$. The number $\phi$ $\beta$ can be taken arbitrarily near to $\pi$.

**3. An extension of Hadamard's theorem.** We shall now consider the case where

$$q(x,u) = (au + bx + c)/(a_1u + b_1x + c_1) \tag{1}$$

and shall proceed in a way analogous to that used by Haslam-Jones* in his extension of Hadamard's theorem to functions of two variables. Suppose that the series

$$\sum_{n=0}^{\infty} a_n u^n, \quad \sum_{n=0}^{\infty} b_n u'^n \tag{2}$$

represent the uniform and analytic functions $f(u)$ and $g(u')$, respectively. Let the singularities $\alpha_i$ of $f(u)$ form an isolated set, $|\alpha_i| \geqq r_1 > 0$. Similarly, let the singularities $\beta_i$ of $g(u')$ form an isolated set, $|\beta_i| \geqq r_2 > 0$. Let $u$ be within a Laurent ring $L(r_0)$ defined by $0 < r_0 < |u| < r_1$, and let $x$ be in a region $R_0$ given by $|x| < r$. Writing $u' = (au+bx+c)/(a_1u+b_1x+c_1)$ we have

$$\begin{aligned} |u'| &= \left|\frac{au + bx + c}{a_1u + b_1x + c_1}\right| = \left|\frac{a + (bx + c)/u}{a_1 + (b_1x + c_1)/u}\right| \\ &< \frac{|a| + (|b|r + |c|)/r_0}{|a_1| - (|b_1|r + |c_1|)/r_0} < r_2 \end{aligned} \tag{3}$$

for $u$ in $L(r_0)$ and $x$ in $R_0$, provided $|a_1| > (|b_1|r + |c_1|)/r_0$ and $|a_1|$ is taken sufficiently great. On the other hand, if $a_1 = 0$ it will be supposed that $c_1 \neq 0$. We have in this case

$$|u'| = \left|\frac{au + bx + c}{b_1x + c_1}\right| < \frac{|ar_1| + |br| + |c|}{|c_1| - |b_1r|} < r_2 \tag{3'}$$

for $u$ in $L(r_0)$ and $x$ in $R_0$, provided $r < |c_1/b_1|$ and $r_2$ is sufficiently great. Let $C$ denote a contour in $L(r_0)$. By (3), the series representing $f(u)$, $g(u')$ are uniformly convergent in $u$ for $u$ on $C$.

Thus $F(x)$ given by

$$F(x) = \frac{1}{2\pi i}\int_C f(u) \cdot g\left(\frac{au + bx + c}{a_1u + b_1x + c_1}\right)\frac{du}{u} = \sum_{n=0}^{\infty} c_n x^n \tag{4}$$

is an analytic function at least in $R_0$. The $c_n$ depend on the $a_i$ and the $b_k$. The integral (4) is analytic in $x$ within a region $R_C$, more extensive than $R_0$, defined by the condition that none of the points in the $u$-plane given by

$$u = \alpha_i, \quad (au + bx + c)/(a_1u + b_1x + c_1) = \beta_k, \quad u = 0, \quad u = \infty \tag{5}$$

* Haslam-Jones, loc. cit., pp. 223–230. Also Faber, loc. cit., p. 282.

shall lie within some assigned positive distance $\delta$ from $C$.* This follows from the fact that for $u$ on $C$ the function $f(u)\cdot g(u')$ is analytic in $x$ for all $x$ in $R_C$, and that it is a continuous function of $x$ and $u$, when $x$ is in $R_C$ and $u$ on $C$. We observe further that, if $C_1$ is a second contour in the $u$-plane and is such that the region between $C$ and $C_1$ contains no points of the set $A$ defined by

$$(6)\quad u=\alpha_i,\qquad (au+bx+c)/(a_1u+b_1x+c_1)=\beta_k,\qquad u=0,\qquad u=\infty,$$

then

$$(7)\quad \int_C f(u)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}\right)\frac{du}{u}=\int_{C_1} f(u)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}\right)\frac{du}{u}.$$

Among the points $A$ it is not necessary to include the point $u$ given by $a_1u+b_1x+c_1=0$. If infinity is included among the values $\beta_i$, this value of $u$ is given by $(au+bx+c)/(a_1u+b_1x+c_1)=\beta_i$ for some $i$. On the other hand, if all the $\beta_i$ are finite, $g(u')$ is regular for $u'=\infty$, that is, $u$ as given by $a_1u+b_1x+c_1=0$ is not a singularity of $g(u')$.

If $x$ is in $R_0(|x|<r)$, $F(x)$ is analytic on account of the convergence of its Taylor's series (4) . This function, however, exists in a more extended region $R_C$ which has with $R_0$ a neighborhood of the origin in common. As $x$ approaches the boundary of $R_C$ some of the singular points in the $u$-plane may approach $C$. To meet this situation $C$ is replaced by $C'$ by a continuous deformation of $C$, without passing over any of the points of the set $A$. The integral over $C_1$ is analytic for $x$ in $R_{C_1}$, that is, in a region including the neighborhood of $x=0$. Thus this integral is an analytic continuation of the power series (4). This process of continuation fails only if points $A$ on opposite sides of $C$ tend to coincide. It follows therefore that $F(x)$ is analytic in every $x$-region which is such that for no point within that region do any two points of the set $A$ coincide. Accordingly, we shall proceed to find values of $x$ for which such coincidences may occur. When a point $u$ defined by $u'=\beta_i$ coincides with a point $\alpha_r$ we have

$$(a\alpha_r+bx+c)/(a_1\alpha_r+b_1x+c_1)=\beta_i.$$

When a point $u$, defined by $u'=\beta_i$, coincides with the point $u=0$, we have $(bx+c)/(b_1x+c_1)=\beta_i$. Coincidences of points $u$, defined by $u'=\beta_i$ with $u=\infty$, would give the relation $a/a_1=\beta_i$ which is contrary to (13); thus, no coincidences of this kind can occur. The coincidence of a $u$-point defined by

* $C$ can now be considered not necessarily restricted to $L(r_0)$; it does not pass through any of the $\alpha_i$ and it contains the origin in its interior.

$u' = \beta_i$ with a $u$-point defined by $u' = \beta_k (k \neq i)$ gives, as a result of solving for $x$ the equation

$$u' - \beta_i = u' - \beta_k = 0,$$

$x = -(c_1a - a_1c)/(b_1a - a_1b)$. Hence the singularities of $F(x)$ are all included among the points

$$(8) \qquad x = \frac{a - a_1\beta_i}{b_1\beta_i - b} \cdot \alpha_r + \frac{c - c_1\beta_i}{b_1\beta_i - b}, \quad x = \frac{c - c_1\beta_i}{b_1\beta_i - b}, \quad x = \frac{-(c_1a - a_1c)}{(b_1a - a_1b)}.$$

We have proved the following theorem:

THEOREM II. *Let* $q(x, u) = (au + bx + c)/(a_1u + b_1x + c_1)$, *where* $|a_1|$ *is sufficiently great so that, for a positive* $r_0$, $r_0 < r_1$, *and a positive* $r$, (3) *is satisfied. If* $a_1 = 0$ *it will be assumed that* $c_1 \neq 0$. *In this case* $r_2$ *is supposed to be sufficiently great so that for a positive* $r$, $r < |c_1/b_1|$, (3′) *is satisfied. Let the series* $\sum a_n x^n$, *with an isolated set of singularities* $\alpha_i$, $|\alpha_i| \geqq r_1$, *and* $\sum b_n x^n$, *with an isolated set of singularities* $\beta_i$, $|\beta_i| \geqq r_2$, *represent the uniform functions* $f(x)$ *and* $g(x)$, *respectively. Let* $F(x)$ *be defined by the integral* (4). *If* $a_1 \neq 0$,

$$(9) \quad F(x) = \sum_n c_n x^n = \sum_{n,m} a_n \cdot b_m \cdot {}_m p_n(x) \qquad \left( q^m(x, u) = \sum_i {}_m p_i(x)/u^i; \, |x| < r \right).$$

*If* $a_1 = 0$,

$$(9') \quad F(x) = \sum_n c_n x^n = a_0 \sum_i b_i s_i(x) \qquad \left( q^m(x, u) = \sum_{i=0}^{i=m} {}_m s_i(x) \cdot u^i; \, |x| < r \right).$$

$F(x)$ *possesses in the whole plane no other singularities than those given by*

$$(10) \qquad \frac{a - a_1\beta_i}{b_1\beta_i - b} \cdot \alpha_r + \frac{c - c_1\beta_i}{b_1\beta_i - b}, \quad \frac{c - c_1\beta_i}{b_1\beta_i - b}, \quad \frac{-(c_1a - a_1c)}{(b_1a - a_1b)}.$$

*There are no singularities for* $|x| < r$.

The second series (9) can be derived as follows. Consider the integral (4). It is an analytic function of $x$ for $x$ in $R_0$ ($|x| < r$). For $u$ in $L(r_0)$, that is, for $0 < r_0 < |u| < r_1$, and $x$ in $R_0$, letting

$$(11) \quad u'^m = \left( \frac{au + bx + c}{a_1u + b_1x + c_1} \right)^m = \sum_i {}_m p_i(x)/u^i, \quad a_1 \neq 0 \quad (\text{here } |u'| < r_2),$$

it is observed that all the series involved in the integrand are absolutely and uniformly convergent. Thus, by multiplication and arranging $f(u) \cdot g(u')$ in powers of $u$, and noting that $F(x)$ is the term free of $u$, we obtain

$$F(x) = \sum_{n,m} a_n b_m \, {}_m p_n(x) \qquad (|x| < r;\ a_1 \neq 0). \tag{12}$$

To derive the second series (9′) we let

$$u'^m = \left(\frac{au + bx + c}{b_1 x + c_1}\right)^m = \sum_{i=0}^{i=m} {}_m s_i(x) \cdot u \qquad (a_1 = 0),$$

where the series converges and $|u'| < r_2$ for $|x| < r$ and $|u| < r_1$. It is observed that $F(x)$ is the term free of $u$ in

$$f(u) \cdot g(u') = \sum_0^\infty d_i(x) \cdot u^i .$$

Hence

$$F(x) = d_0(x) = a_0 \cdot \sum_{i=0}^{\infty} b_i \cdot {}_i s_0(x) \qquad (|x| < r;\ a_1 = 0). \tag{12′}$$

We see therefore that in the case when $a_1 = 0$ the singularities of the function $F(x)$ are independent of the singularities of $f(x)$. This case is consequently of no special interest. A priori it would seem to be possible to get Hurwitz's theorem* as a particular case of Theorem II by setting $a_1 = b_1 = c = 0$, $b = c_1 = 1$, $a = 1$. Indeed introducing these values in formula (10) we get the singularities of the composed function of Hurwitz's theorem. Unfortunately, however, this function is not represented by (9′) which is an entirely different function. While the integrals look alike, the contours of integration are different. It is not possible to choose the contour in Theorem II in such a fashion that both Hadamard's and Hurwitz's theorems come out as special cases.

If in Theorem II we let $b = a_1 = 1$ and $a = c = b_1 = c_1 = 0$, $q(x, u)$ becomes $x/u$ and (3) becomes $r/r_0 < r_2$. The latter inequality is seen to be satisfied for suitable values of $r$ and $r_0$. The function $F(x)$, as given by (9), will become $F(x) = \sum a_n b_n x^n$. On the other hand, since there are no singularities of $F(x)$ for $|x| < r$, we have from (10) the expressions $\beta_i \alpha_r$ as the only ones among which all of the singularities of $F(x)$ are found. Thus Theorem II contains Hadamard's theorem, as a particular case.

4. **The extension of the theorem of §3 to functions of two variables.**† Let $f(u, v)$ be an analytic function of the complex variables $u$ and $v$, having a branch regular near $u = 0, v = 0$. Let for that branch

$$f(u,v) = \sum \sum a_{mn} u^m v^n . \tag{1}$$

* S. Mandelbrojt, loc. cit., p. 24.

† Compare with the methods used by Haslam-Jones, loc. cit., pp. 223–230.

Similarly, let $g(u', v')$ be another function for which

$$g(u', v') = \sum\sum b_{mn} u'^m v'^n. \tag{2}$$

Suppose that the singularities of $f(u, v)$ and $g(u', v')$ are given by $p(u, v) = 0$ and $q(u', v') = 0$, respectively. The functions $p(u, v)$ and $q(u', v')$ are analytic functions of type $Z$. *A function $s(u, v)$ is said to be of type Z if for any constant $u_1$, $s(u_1, v) \equiv 0$ or has isolated zeros in the v-plane; and for any constant $v_1$, $s(u, v_1) \equiv 0$ or has isolated zeros in the u-plane.** Let $f_1, f_2$ and $g_1, g_2$ be positive constants such that (1) is convergent for $|u| < f_1$, $|v| < f_2$, and (2) is convergent for $|u'| < g_1$, $|v'| < g_2$. Then for $|a_1|$ sufficiently large, positive constants $u_1$ and $v_1$ can be found, $u_1 < f_1$, $v_1 < f_2$, such that

$$\left|\frac{au + bx + c}{a_1u + b_1x + c_1}\right| < g_1, \qquad \left|\frac{av + by + c}{a_1v + b_1y + c_1}\right| < g_2 \tag{3}$$

for all $u$ and $v$ satisfying $|u| > u_1$, $|v| > v_1$ and for all $x$ and $y$ in a region $R_0(|x| < r, |y| < s)$. As in (3) in §3, writing $u' = (au+bx+c)/(a_1u+b_1x+c_1)$ and $v' = (av+by+c)/(a_1v+b_1y+c_1)$, we have both inequalities

$$|u'| = \left|\frac{a + (bx + c)/u}{a_1 + (b_1x + c_1)/u}\right| < \frac{|a| + (|b|r + |c|)/u_1}{|a_1| - (|b_1|r + |c_1|)/u_1} < g_1, \tag{4}$$

$$|v'| = \left|\frac{a + (by + c)/v}{a_1 + (b_1x + c_1)/v}\right| < \frac{|a| + (|b|s + |c|)/v_1}{|a_1| - (|b_1|s + |c_1|)/v_1} < g_2 \tag{5}$$

satisfied, provided $|a_1|$ is sufficiently great. Let $u$, $v$ be within the region $L$ defined by the relations $u_1 < |u| < f_1$, $v_1 < |v| < f_2$. In $L$, both $f(u, v)$ and $g(u', v')$ have their respective Taylor's series absolutely and uniformly convergent for all $x$, $y$ in $R_0$. We shall consider now the following integral:

$$F(x, y) = \frac{-1}{4\pi^2}\int_C \int_{C'} f(u,v) \cdot g\left(\frac{au + bx + c}{a_1u + b_1x + c_1}, \frac{av + by + c}{a_1v + b_1y + c_1}\right)\frac{dv}{v}\cdot\frac{du}{u}, \tag{6}$$

where $C$ is a contour in the $u$-plane and $C'$ a contour in the $v$-plane, both in $L$. Substituting the power series (1) and (2) in (6), these series being absolutely and uniformly convergent on $C$ and $C'$, it is observed that multiplication term by term is possible. Thus for $x$, $y$ in $R_0$

$$F(x, y) = \sum\sum c_{mn} \cdot x^m y^n \tag{7}$$

where the $c_{mn}$ depend on the $a_{mn}$ and the $b_{mn}$.

The integral (6) is analytic for $(x, y)$ in $R_0$ ($|x| < r$, $|y| < s$). Using (11) of §3, noting the absolute and uniform convergence of all of the series in-

* For this definition see Haslam-Jones, loc. cit., p. 223.

volved and arranging $f(u, v)$, $g(u', v')$ in powers of $v$ ($u$ and $v$ in $L$), we get from the product

$$\sum_s v^s\Big(\sum_i a_{is}u^i\Big)\cdot\sum_s v^{-s}\Big(\sum_{m,k,r} b_{mk}\,{}_mp_r(x)\,{}_kp_s(y)/u^r\Big)$$

the term free of $v$

$$\frac{1}{2\pi i}\int_{C_1} f\cdot g\cdot\frac{dv}{v} = \sum_s\Big[\Big(\sum_r a_{rs}u^r\Big)\cdot\Big(\sum_r B_{r,s}(x,y)u^{-r}\Big)\Big], \tag{8}$$

where

$$B_{r,s}(x,y) = \sum_{m,k} b_{mk}\,{}_mp_r(x)\,{}_kp_s(y). \tag{9}$$

Multiplying (8) by $du/(2\pi iu)$, and integrating with respect to $u$ along the contour $C$, and arranging (8) in powers of $u$, we obtain the term free of $u$

$$F(x,y) = \sum_{r,s} a_{rs}B_{r,s}(x,y) = \sum_{r,s,m,k} a_{rs}b_{mk}\,{}_mp_r(x)\,{}_kp_s(y). \tag{10}$$

The series in the last member converges at least in $R_0$.

We shall now proceed to find all of the possible singular loci of the function $F(x, y)$ whose analytic element in $R_0$ is given by (7). Denote the values of $v$ satisfying $p(u, v)=0$ by $(\Gamma u)$. Let the $d_i$ be the points of the $u$-plane for which $p(u, v)\equiv 0$ for all $v$, and which are therefore defined by the equation $p(u, 0)=0$. The $d_i$ form an isolated set. Denote the values of $v$ satisfying $q((au+bx+c)/(a_1u+b_1x+c_1),\ (av+by+c)/(a_1v+b_1y+c_1))=0$ by $(\Gamma u')$ and the isolated values of $u$ satisfying $q(\ (au+bx+c)/(a_1u+b_1x+c_1),\ 0)=0$ by $\phi_i$. The $(\Gamma' u)$ depend on $u$, $x$ and $y$; the $\phi_i$ depend on $x$. We shall now show that *if $(x, y)$ is in $R_0$ and if $u$ describes $C$ ($C$ and $C'$ in $L$), the points $v=(\Gamma' u)$ describe contours $\Gamma$ outside of $C'$ and the points $v=(\Gamma' u)$ describe contours $\Gamma'$ inside $C'$, provided $|a'|$ is large enough.* For $u$ on $C$, $|u|<f_1$. If $|v|=|(\Gamma u)|<f_2$, there would be a contradiction to the assumption that (1) converges whenever $|u|<f_1$, $|v|<f_2$. Hence $|(\Gamma u)|\geqq f_2$, and thus the $\Gamma$ contours are outside of $C'$. Similarly, when $x$ and $y$ are in $R_0$ the values $v'=(G'u')$ defined by $q(u', v')=0$ satisfy the inequality $|v'|\geqq g_2$. The points $v=(\Gamma' u)$ satisfy the equation $(av+by+c)/(a_1v+b_1y+c_1)=v'=(G'u')$. Thus

$$v = (\Gamma' u) = \frac{(b_1y+c_1)v' - (by+c)}{a - a_1v'} \quad \text{where } |v'|\geqq g_2. \tag{11}$$

We have

$$|(\Gamma' u)| = \left|\frac{(b_1y+c_1) - (by+c)/v'}{a_1 - a/v'}\right| < \frac{|b_1|s+|c_1|+(|b|s+|c|)/g_2}{|a_1|-|a|/g_2} \leqq v_1, \tag{12}$$

provided that $|a_1|$ is sufficiently large. The constant $|a_1|$ will be taken so that the inequalities (4), (5) and (12) are all satisfied. Having so chosen $|a_1|$, from (12) it follows that the contours $\Gamma'$ are all inside $C'$. Thus, for $C$ in $L$ and $(x, y)$ in $R_0$, the contour $C'$ in the $v$-plane separates the $\Gamma$ figure (consisting of the contours $\Gamma$) from the $\Gamma'$ figure (consisting of the contours $\Gamma'$).

Introduce in the $v$-plane a contour $C_u$ depending on $u$ and consisting of a circle of a large radius $R$, $R>|(\Gamma' u)|$, with the center $v=0$. Let it be indented to exclude those points $v=(\Gamma u)$ which are inside of this circle, and let it contain all the points $(\Gamma' u)$. (In order that this should be possible, for $(x, y)$ not necessarily in $R_0$ and $u$ not necessarily in $L$, the conditions $(\Gamma' u)\neq\infty$, $(\Gamma u)\neq(\Gamma' u)$ will be later introduced, the first to have a finite $R>|(\Gamma' u)|$, the second to make separation of the points $(\Gamma u)$ from the points $(\Gamma' u)$ possible.) When $u$ is on $C$ and $(x, y)$ in $R_0$ the integrand of

$$\int_{C'} f(u,v)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right)\cdot\frac{dv}{v}, \tag{13}$$

as a function of $v$, has no singular points in the region of the $v$-plane bounded by $C'$ and $C_u$. Accordingly,

$$\int_{C'} f(u,v)\cdot g(u',v')dv/v = \int_{C_u} f(u,v)\cdot g(u',v')dv/v \tag{14}$$

and

$$F(x,y) = -\frac{1}{4\pi^2}\int_C\int_{C_u} f(u,v)\cdot g(u',v')\cdot\frac{dv}{v}\cdot\frac{du}{u} \tag{15}$$

which holds for $(x, y)$ in $R_0$ and $C$ in $L$. Concerning this integral we shall prove the following lemma:

LEMMA II. *With $C$ any contour in the $u$-plane, the integral* (15) *is analytic in $x$ and $y$ for $(x, y)$ in a region $R_C$ defined by the condition that none of the points of the $u$-plane given by*

$$p(u,v) = q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right) = 0, \tag{16}$$

*or*

$$q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{a}{a_1}\right) = 0^* \tag{17}$$

*shall lie within an assigned distance $\delta$, $\delta>0$, from $C$.*

* (17) is obtained from $q(u', v')=0$ by letting $v\to\infty$.

When $u$ is a fixed point on $C$, $(\Gamma' u) \neq (\Gamma u)$ by (16); and $(\Gamma' u) \neq \infty$ by (17). Thus the contour $C_u$ can be constructed with a finite $R$, so that it separates the points $(\Gamma' u)$ and $(\Gamma u)$ in such a way that these points do not lie within some assigned positive distance from $C_u$. Consequently, we can have a contour $\overline{C}_u$ which is such that $C_u$ can be continuously deformed into $\overline{C}_u$, without encountering any of the $v$-singularities of the integrand. Then

$$\int_{C_u} f(u,v)\cdot g(u',v')dv/v = \int_{\overline{C}_u} f(u,v)\cdot g(u',v')dv/v, \tag{18}$$

where the second integral (18) remains analytic in $(x, y)$. Thus, *for $u$ fixed on $C$, and $(x, y)$ in $R_C$, the first integral* (18) *is analytic in $x$ and $y$.* We shall now prove that this integral is analytic in $x$, $y$ and $u$, for $(x, y)$ in $R_C$ and $u$ within a small circle $s$ with center at $u$, $u$ a point on $C$; that is, that *it is continuous in $x$, $y$ and $u$ for $(x, y)$ in $R_C$ and $u$ on $C$.* This will complete the proof of the lemma. Choose $s$ small enough so that the representations of $s$ on the $\Gamma$ and $\Gamma'$ figures, $\sigma$ and $\sigma'$ respectively, have no points in common. If, for $s$ no matter how small, $\sigma$ and $\sigma'$ had a point in common, (13) would hold for $u$ on $C$ and $(x, y)$ in $R_C$. Let $C_\sigma$ be a contour excluding $\sigma$ and including $\sigma'$. Then

$$\int_{C_u} f(u,v)\cdot g(u',v')dv/v = \int_{C_\sigma} f(u,v)\cdot g(u',v')dv/v. \tag{19}$$

As in the case of the second integral (18), the second integral (19) is analytic in $x$, $y$ and $u$ when $(x, y)$ is in $R_C$ and $u$ is in $s$. Hence the same holds true of the first integral (19).

LEMMA III. *Given in the $u$-plane a contour $C_1$ such that $C$ can be deformed into $C_1$ by a continuous deformation without passing over any of the points $A$ defined by*

$$p(u,v) = q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right) = 0, \tag{20}$$

$$q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{a}{a_1}\right) = 0, \tag{21}$$

$$p(u,0) = 0, \tag{22}$$

$$u = 0, \quad u = \infty, \tag{23}$$

*then*

$$\int_C \int_{C_u} f(u,v)\cdot g(u',v')\cdot\frac{dv}{v}\cdot\frac{du}{u} = \int_{C_1}\int_{C_u} f(u,v)\cdot g(u',v')\cdot\frac{dv}{v}\cdot\frac{du}{u}. \tag{24}$$

We denote by $D$ the region in the $u$-plane bounded by $C$ and $C_1$, the representations of $D$ on the $\Gamma$ and $\Gamma'$ figures by $\Delta$ and $\Delta'$, respectively, and the contour in the $v$-plane containing $\Delta'$ and excluding $\Delta$ by $C_\Delta$. If $\Delta$ and $\Delta'$ have no points in common

$$\frac{1}{u}\int_{C_u} f(u,v)\cdot g(u',v')\cdot\frac{dv}{v} = \frac{1}{u}\int_{C_\Delta} f(u,v)\cdot g(u',v')\cdot\frac{dv}{v}\cdot \tag{25}$$

For $(x, y)$ fixed, the second, and therefore the first, integral is an analytic function of $u$ for all $u$ in $D$. When $\Delta$ and $\Delta'$ have points in common we may use precisely the procedure of Haslam-Jones.* Thus, the function of $u$, given by the first integral (25), can be integrated along $C$ giving the first integral (24). $C$ can be then deformed into $C_1$. This completes the proof of the lemma.

$F(x, y)$, which is analytic in $R_0$, is given there by the convergent power series (7). By Lemma II, $F(x, y)$ is analytic in the more extensive region $R_C$, its analytic expression there being given by (15). The region of validity of the representation of $F(x, y)$, by means of a double integral, can be further extended by deforming $C$ into $C_1$, utilizing Lemma III. It is impossible to continue this process when two points of the set $A$, lying on opposite sides of $C$ (the contour in the $u$-plane), coincide. Thus, $F(x, y)$ has a principal branch all of whose singularities are found among the relations expressing coincidence of two points of the set $A$. This set is defined by (20), (21), (22), (23). Using Hartogs' theorem that an analytic function of two variables cannot have isolated points for singularities, and a lemma by Haslam-Jones,† coincidences of the points of the set $A$ will be shown to yield no singularities of $F(x, y)$. Coincidences of two solutions of (21), of a solution of (21) with that of (22), of a solution of (22) with that of (23), may occur only for isolated values of $x$; the points $u$ defined by those equations being independent of $y$, there arise no singularities. There are no singularities of $F(x, y)$ in connection with (22), (23).

---

* Haslam-Jones, loc. cit., p. 227.

† Hartogs, Münchener Sitzungsberichte, vol. 36 (1906), p. 223. Haslam-Jones, loc. cit., p. 228. We quote the lemma:

*Let $f(u, x, y)$ be an analytic function of $u$, $x$, and $y$ which has continuous singularities $u=\alpha_r(x, y)$, and consider a contour $C$ in the $u$-plane. If, for $\phi(x, y)=0$, $\alpha_1=\alpha_2$ and the other singularities are such that $|\alpha_r-\alpha_s|>\delta>0$, then*

$$F(x, y) = \int_C f(u, x, y)du$$

*is regular for $\phi(x, y)=0$ provided that $f(u, x, y)=f_1(u, x, y)+f_2(u, x, y)$, where $f_1(u, x, y)$ is regular for $u=\alpha_1(x, y)$, and $f_2(u, x, y)$ is regular for $u=\alpha_2(x, y)$.*

Since the case when $F(x, y)$ is a constant is of no interest, it will be supposed that $b$ and $b_1$ are not zero together and, if $a=0$, $b=0$, then $c\neq 0$ If $a=0$ and $b\neq 0$ the values of $x$ and $y$ satisfying $bx+c=by+c=0$ will, in the following, be excluded. If $a\neq 0$, it will be assumed that $bc_1-b_1c\neq 0$.

Suppose that a point $u=\alpha_2(x, y)$, satisfying (20), coincides with a point $u=\alpha_1$, satisfying (22). Suppose $|v|\to 0$, as $\alpha_2(x, y)\to\alpha_1$. For $v=0$, (20) becomes

$$p(u,0) = q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{by+c}{b_1y+c_1}\right) = 0. \tag{26}$$

We have the isolated set of values $u$ defined by $p(u, 0)=0$, $u=d_r$, and from the second member, $u=e_s(x, y)$. Therefore singularities may arise for $d_r=e_s(x, y)$, that is, for $x=n_s(d_r, y)$. Substituting this expression for $x$ in (20), eliminating $v$, and letting $u=d_r$ we derive an isolated set of values $y$, $y=m_i$, and hence an isolated set of values $x$, $x=n_s(d_r, m_i)$. By Hartogs' theorem no singularities, due to this case, arise. Suppose $|v|$ does not approach 0, as $\alpha_2(x, y)\to\alpha_1$. In (20), let $u=\alpha_2(x, y)=\alpha_1$. Assume that whenever $|v|<\epsilon$,

$$|p(\alpha_1, v)| \leqq \eta, \tag{27}$$

$$\left|q\left(\frac{a\alpha_1+bx+c}{a_1\alpha_1+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right)\right| \leqq \eta \qquad (\eta\to 0, \text{ as } \epsilon\to 0). \tag{28}$$

The two inequalities cannot hold simultaneously, since that would mean that $v\to 0$, as $\alpha_2\to\alpha_1$. The inequality (27) holds since $p(\alpha_1, 0)=0$. Therefore

$$\left|q\left(\frac{a\alpha_1+bx+c}{a_1\alpha_1+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right)\right| > \eta, \quad \text{whenever} \quad |v|<\epsilon. \tag{29}$$

Let $C_{u0}$ be a contour in the $v$-plane consisting of $C_u$ and in addition indented to exclude the origin. We have then

$$\begin{aligned}\int_{C_u} f(u,v)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right)\cdot\frac{dv}{v} \\ = \int_{C_{u0}} f(u,v)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right)\cdot\frac{dv}{v} \\ + \frac{1}{2\pi i}\cdot f(u,0)\cdot g\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{by+c}{b_1y+c}\right).\end{aligned} \tag{30}$$

This relation follows from the following considerations. Restrict $(x, y)$ so that none of the $u$-solutions of (20) coincide with a solution of (22). Then

$p(\alpha_2(x, y), 0) \neq 0$, and $f(u, v)$ is thus analytic for $|u-\alpha_2|<\delta$, $|v|<\epsilon$ (for $\delta$, $\epsilon$ sufficiently small). The function $g(u', v')$ has $u=\alpha_2(x, y)$, $v=0$ for singular values only if

$$q\left(\frac{a\alpha_2(x,y)+bx+c}{a_1\alpha_2(x,y)+b_1x+c_1}, \frac{by+c}{b_1y+c_1}\right)=0.$$

If this relation were true, one of the equations (20), $g(u', v')=0$, where $u$ is replaced by $\alpha_2(x, y)$, would yield a solution $v=0$. Another equation (20), $p(\alpha_2(x, y), v)=0$, would become $p(\alpha_2(x, y), 0)=0$, and this was shown to be impossible. Thus, $g(u', v')$ is analytic for $|u-\alpha_2|<\delta$, $|v|<\epsilon$, and the same is true of $f(u, v)\cdot g(u', v')$. Hence $f(u, v)\cdot g(u', v')/(2\pi i)$, with $v=0$, is a function of $u$ analytic in the neighborhood of $u=\alpha_2(x, y)$. This function is the last term in (30) and is the residue at $v=0$ of the integrand in (30). By (29), $g(u', v')$ is analytic for $|u-\alpha_1|<\delta$, $|v|<\epsilon$ and hence the integral in the second member of (30) is a function of $u$ analytic in a neighborhood of $u=\alpha_1$. Hence, by the lemma of Haslam-Jones, there are no singularities due to the case under consideration (with the above hypothesis on $x$ and $y$).

If $a=0$, $b\neq 0$, the preceding reasoning is neither applicable for $x=-c/b$ nor for $y=-c/b$. Consider the case $a=0$, $b\neq 0$, $y=-c/b$. With $a=0$, $b\neq 0$, from (11) of §3 it can be shown that every ${}_mp_{m+i}(x)$ contains $(bx+c)$ to the $m$th power, as a factor, and that ${}_mp_r(x)\equiv 0$ for $r<m$. Thus letting $y=-c/b$ in (9), and observing that ${}_kp_s(-c/b)=0$, unless $k=s=0$ $({}_0p_0(-c/b)=1)$, we get

$$B_{r,s}(x, -c/b)=0, \quad s\geqq 1,$$

and

$$B_{r,0}(x, -c/b)=\sum_m b_{m,0}\cdot {}_mp_r(x).$$

Thus from (10)

$$F(x, -c/b)=\sum_r a_{r,0}\cdot B_{r,0}(x, -c/b)=\sum_{r,m} a_{r,0}\cdot b_{m,0}\cdot {}_mp_r(x). \tag{31}$$

Since

$$f(u,0)=\sum_r a_{r,0}\cdot u^r, \qquad g(u,0)=\sum_m b_{m,0}\cdot u^r,$$

by comparing (31) with (12) of §3, it can be seen that $F(x, -c/b)$, as given by the methods of this section, is derivable from $f(u, 0)$, $g(u, 0)$ by means of Theorem II. This theorem is applicable since, by the hypothesis on $p(u, v)$, $q(u, v)$, the functions $f(u, 0)$, $g(u, 0)$ have isolated singularities. Thus $F(x, -c/b)$ has only isolated singularities. Similarly, $F(-c/b, y)$ has only isolated singularities. By Hartogs' theorem no singularities of $F(x, y)$ arise for this case.

Singularities, due to the coincidences of a $u$-point of (20) with a point of (21) or a point of (23), are eliminated in the particular case treated by Haslam-Jones, by methods which cannot be readily extended to our case. Neither did we succeed in eliminating these singularities by any other method. Thus we assert the following.

All the singular loci of $F(x, y)$ are found among the relations expressing coincidence of two $u$-points of (20), of a point of (20) with a point of (21), of a point of (20) with that of (23). The points (23) are $u=0, u=\infty$.

We have thus proved the following theorem:

THEOREM III. *Let $f(u,v)$ be an analytic function of $u$ and $v$ having a branch regular at $u=0, v=0$, and having for that branch*

$$f(u,v) = \sum\sum a_{m,n}u^m v^n \quad (\textit{convergent for } |u|<f_1,\ |v|<f_2); \tag{32}$$

*let $g(u', v')$ be a similar function for which*

$$g(u',v') = \sum\sum b_{m,n}u'^m v'^n \quad (\textit{convergent for } |u'|<g_1,\ |v'|<g_2), \tag{33}$$

*and let $f(u, v)$ and $g(u', v')$ have singularities given by $p(u, v)=0$ and $q(u', v')=0$, respectively, where $p(u, v)$ and $q(u, v)$ are analytic functions of the type Z. Let $r(x, u)=(au+bx+c)/(a_1u+b_1x+c_1)$, where $|a_1|$ is sufficiently great, so that for positive $r$, $s$, $u_1$, $v_1(u_1<f_1, v_1<f_2)$ the inequalities (4), (5) and (12) are all satisfied. Exclude the relation $b=b_1=0$. If $a\neq 0$, let $bc_1-b_1c\neq 0$.*

*The function $F(x, y)$, given by the series*

$$F(x,y) = \sum_{m,n} c_{m,n}x^m y^n = \sum_{r,s,m,k} a_{r,s}\cdot b_{m,k}\cdot {}_mp_r(x)\cdot {}_kp_s(y) \tag{34}$$

$$\left(r^m(x,u) = \sum_i {}_mp_i(x)/u^i\right)$$

*converging for $|x|<r$, $|y|<s$, has a branch which is regular near $x=0$, $y=0$. It has all of its singular loci included among the $(x, y)$-relations expressing coincidence of two $u$-solutions of*

$$p(u,v) = q\left(\frac{au+bx+c}{a_1u+b_1x+c_1}, \frac{av+by+c}{a_1v+b_1y+c_1}\right) = 0, \tag{20}$$

*of a solution of (20) with that of $p(u, 0)=0$, of a solution of (20) with $u=0$, of a solution of (20) with $u=\infty$.*

5. **An extension of Theorem II.** Consider the integral

$$\frac{1}{2\pi i}\int f(u)g(q(x,u))du/u,$$

where

$$q(x,u) = \frac{e_0 + e_1u + \cdots + e_nu^n}{g_0 + g_1u + \cdots + g_nu^n}, \tag{1}$$

$$\begin{aligned} e_i &= e_{i0} + e_{i1}x + \cdots + e_{im}x^m, \\ g_i &= g_{i0} + g_{i1}x + \cdots + g_{im}x^m. \end{aligned} \tag{2}$$

The constants $e_{ik}$, $g_{ik}$, with the exception of $g_{n0}$, satisfy the inequalities

$$|e_{ik}|, \quad |g_{ik}| \leqq s.$$

Also $e_n \not\equiv 0$, $g_n \not\equiv 0$. Concerning the functions $f(u), g(u')$ we make the assumptions of §3. For $u$ within a Laurent ring $L(r_0)$, defined by $0 < r_0 < |u| < r_1$, and $x$ within a region $R_0$, defined by $|x| < r$, the following holds:

$$|g_i(x)|, |e_i(x)| < s\cdot(1 + r + \cdots + r^m) = s_m(r). \tag{3}$$

The inequalities (3) do not include $g_n(x)$; however, $|g_n(x) - g_{n0}| < s_m(r)$. Also,

$$\begin{aligned} |g_n + g_{n-1}u^{-1} + \cdots + g_0u^{-n}| &= |g_{n0} + (g_n - g_{n0}) + g_{n-1}u^{-1} \\ &\quad + \cdots + g_0u^{-n}| > |g_{n0}| - s_m(r)(1 + r_0^{-1} + \cdots + r_0^{-n}) > 0, \end{aligned}$$

provided $|g_{n0}|$ is sufficiently great. Thus writing $u' = q(x, u)$, it follows that

$$\begin{aligned} |u'| = |q(x,u)| &= \left|\frac{e_n + e_{n-1}u^{-1} + e_{n-2}u^{-2} + \cdots + e_0u^{-n}}{g_n + g_{n-1}u^{-1} + g_{n-2}u^{-2} + \cdots + g_0u^{-n}}\right| \\ &< \frac{s_m(r)\cdot(1 + r_0^{-1} + r_0^{-2} + \cdots + r_0^{-n})}{|g_{n0}| - s_m(r)\cdot(1 + r_0^{-1} + \cdots + r_0^{-n})} < r_2, \end{aligned} \tag{4}$$

provided that $|g_{n0}|$ is sufficiently great and $x$ is in $R_0$ and $u$ in $L(r_0)$.

Let $C$ denote a closed contour in $L(r_0)$. The series representing $f(u)$ and the series representing $g(u')$ converge absolutely and uniformly when $x$ is in $R_0$ and $u$ on $C$. The latter series converges since, by (4), $|u'| < r_2$. Thus $F(x)$, given by

$$\begin{aligned} F(x) &= \frac{1}{2\pi i}\int_C (\sum a_nu^n)\cdot(\sum b_nq^n(x,u))\cdot du/u \\ &= \sum c_nx^n = \sum_{n,m} a_n\cdot b_m\cdot {}_ms_n(x), \end{aligned} \tag{5}$$

is an analytic function. The last two representations (5) converge at least in $R_0$. The second of the two representations is derived by noting that

$$q^m(x,u) = \sum_i {}_ms_i(x)/u^i \tag{6}$$

is absolutely and uniformly convergent for $x$ in $R_0$ and $u$ in $L(r_0)$, and by proceeding in the way similar to the one used in deriving (12) of §3.

The integral (5) is analytic within a wider region $R_C$, defined by the condition that none of the points $A$ in the $u$-plane, given by

$$u = \alpha_i, \quad q(x,u) = \beta_k, \quad u = 0, \quad u = \infty \ , \tag{7}$$

shall lie within some assigned positive distance from $C$. The contour $C$ is now not necessarily in $L(r_0)$, it contains $u=0$ in its interior, and does not pass through any of the $\alpha_i$. Analyticity of (5) follows from analyticity of $f(u)\cdot g(q(x,u))$ in $x$, for all $x$ in $R_C$, and $u$ any fixed point on $C$; and from the continuity of this product in $x$ and $u$, for $x$ in $R_C$ and $u$ on $C$. Further, if $C_1$ is a second contour in the $u$-plane containing $C$ and there are no points of the set $A$ in the region bounded by $C$ and $C_1$ (nor on $C_1$), then

$$\int_C f(u)\cdot g(u')\cdot du/u = \int_{C_1} f(u)\cdot g(u')\cdot du/u. \tag{8}$$

Among the points $A$ it is not necessary to include the points $u$ defined by

$$g_0 + g_1\cdot u + \cdots + g_n u^n = 0. \tag{9}$$

Any such point gives $q(x,u)\equiv u' = \infty$. If infinity is included among the values $\beta_i$, it has already been given by $q(x,u)=\beta_i$, for some $i$. If the $\beta_i$ are bounded as a set, $g(u')$ is regular for $u'=\infty$, that is, it is regular for any $u$ satisfying (9). By repeating the reasoning of §3, it is found that all of the singularities of $F(x)$ are included among the values $x$ for which two of the points of the set $A$ coincide. Coincidences of a $u$-solution of $q(x,u)=\beta_k$ with $u=\alpha_i$, $u=0$, $u=\infty$ give

$$q(x,\alpha_i) = \beta_k, \qquad q(x,0) = 0, \qquad q(x,\infty) = 0, \tag{10}$$

respectively. Coincidences of two $u$-solutions, whether they do belong to the same equation of the set of equations $q(x,u)=\beta_k$, or not, give also possible singularities of $F(x)$. We are able thus to state the following theorem:

THEOREM IV. *Let $q(x,u)$ be a function defined by* (1) *and* (2). *Let $|g_{n0}|$ be sufficiently great so that, for a positive $r_0$, $r_0<r_1$, and a positive $r$,* (4) *is satisfied. Let the series $\sum a_n x^n$ with an isolated set of singularities $\alpha_i$, $|\alpha_i|\geqq r_1$, and $\sum b_n x^n$ with an isolated set of singularities $\beta_i$, $|\beta_i|\geqq r_2$, represent the uniform functions $f(x)$ and $g(x)$, respectively. The function*

$$F(x) = \sum_n c_n x^n = \sum_{n,m} a_n\cdot b_m\cdot {}_m s_n(x), \qquad q^m(x,u) = \sum_i {}_m s_i(x)/u^i,$$

*where both series converge at least for all* $|x|<r$, *possesses in the whole plane singularities which are all included among the values of* $x$ *satisfying*

$$q(x, \alpha_i)=\beta_k, \quad q(x, 0)=0, \quad q(x, \infty)=0$$

*and the values of* $x$ *for which two* $u$*-solutions of* $q(x, u)=\beta_k$ *coincide.*

LEHIGH UNIVERSITY,
BETHLEHEM, PA.

# RESIDUES OF POLYGENIC FUNCTIONS*

BY

VINCENT C. POOR

1. **Important results.** A function $W(z)=u+iv$ is called polygenic if it is not analytic or monogenic. The delay in the study of such functions has doubtless been due to the fruitful field of analytic functions opened up by the uniqueness restriction on the derivative, $dw/dz$, of $w$. Some of the most important results in this relatively new field of polygenic functions are the areal derivative

$$\frac{dw}{d\alpha} = \lim_{A\to 0} \frac{i\int_c wdz}{2A}$$

introduced with a different factor by Pompeiu† and

$$\frac{dw}{d\beta} = \lim_{A\to 0} \frac{-\int_c wd\bar{z}}{2iA}$$

which in its essence is Kasner's‡ mean derivative.

If $W$ is analytic in $z$ or $\bar{z}$, one of these circulation integrals is identically zero. Also from these two definitions the two circulation theorems§

$$2\int_A \frac{dw}{d\alpha}dA = -i\int_c wdz; \quad 2i\int_A \frac{dw}{d\beta}dA = -\int_c wd\bar{z}$$

are almost evident.

We define a polygenic function as regular at a point when it possesses a differential at that point. In the sequel we shall assume that the functions we consider are, in general, regular so that for a regular point, $a$, of a function $f(z)$ we will have

---

* Presented to the Society, November 30, 1929; received by the editors in September, 1929.

† Pompéiu, *Sur une classe de fonctions d'une variable complexe*, Rendiconti del Circolo Matematico di Palermo, vol. 23 (1st semester 1912), pp. 108–113.

‡ Kasner, *General theory of polygenic or non-monogenic functions. The derivative congruence of circles*, Proceedings of the National Academy of Sciences, January, 1928, pp. 75–82. The notation here is due to G. Y. Rainich, University of Michigan.

§ Calugaréano (Thesis), *Sur les fonctions polygènes d'une variable complexe*, Paris, 1928, pp. 11, 12.

$$(1.1) \qquad f(z) - f(a) = \frac{df}{d\beta}(z - a) + \frac{df}{d\alpha}(\bar{z} - \bar{a}) + \eta$$

where $df/d\beta$ and $df/d\alpha$ are the values of the derivatives at the point $a$, and $\eta$ is a quantity of higher order than the first in $z-a$, a quantity such that $\eta/(z-a)^2$ remains finite when $z$ goes to $a$.

In studying such functions we have as our model the highly developed theory of analytic or monogenic functions.

2. **Zeros and poles.** The simplest singularities of monogenic functions are poles which may be characterized as the zeros of the reciprocal of a given function. We begin therefore by a brief discussion of zeros of polygenic functions.

There is a large class of polygenic functions defined by a double power series*

$$f(z) = a_0 + a_{10}z + a_{01}\bar{z} + \cdots .$$

Such a function vanishes at the origin if the absolute term, $a_0$, is taken equal to zero. In the case of a monogenic function the result of such vanishing is that a factor $z$ may be taken out so that the function can be represented in the form

$$f(z) = zF(z)$$

where $F(z)$ is regular at the origin. A polygenic function of this form, where $F(z)$ is another polygenic function, also has a zero at the origin but this is by no means the only type of zeros. Without going far into the classification of zeros we consider in addition to this what seems to be the next simplest case, the case in which the whole linear term may be taken out as a factor so that

$$f(z) = (az + b\bar{z})F(z)$$

and we assume that $F(z)$ does not vanish in the neighborhood of the origin.

The function $az+b\bar{z}$ evidently vanishes for $z=0$; to see whether it vanishes for other points in the neighborhood of the origin we notice that if $az+b\bar{z}$ is zero we must have

$$az = -b\bar{z}$$

or taking absolute values

$$|a| = |b| .$$

Thus if $|a| \neq |b|$ the origin is an isolated zero of $f(z)$. But if the relation

* Calugaréano, loc. cit., § 1.

$|a| = |b|$ or its equivalent $b = ae^{i\phi}$ holds we easily see that we have a whole line of zeros, distributed along the straight line whose slope angle is $(\phi - \pi)/2$.

We pass to the consideration of poles which we define in accordance with the above, as points at which the function becomes infinite while its reciprocal is zero. Corresponding to the two simple types of zeros mentioned we consider two types of poles of $f(z)$ given by the equations

$$f(z) = \frac{F(z)}{z} \text{ and } f(z) = \frac{F(z)}{az + b\bar{z}}$$

respectively; here the pole is assumed, for convenience in writing, to be at the origin.

3. **Residues.** In the theory of *monogenic* functions the residue is defined by the equation

$$R_A = \frac{1}{2\pi i}\int_c f(z)dz.$$

If $f(z)$ has no poles inside the closed contour $c$, defining the region $A$, then $R_A$ is zero. If in this region $f(z)$ has a pole so that

$$F(z) = (z - a)f(z),$$

then

$$R_A = F(a).$$

The only restriction on the contour $c$ is that the region $A$ bounded by it contains no other singularities of $f(z)$.

If, on the other hand, $f(z)$ is a *polygenic* function the contour integral is no longer independent of the path. Therefore the value of $R_A$ given above depends on the contour $c$, and the definition loses its sense. We may, however, *define* the residue of a polygenic function as the limit toward which the above expression tends when the contour $c$, assumed to be a circle* of radius $r$ with center at $a$, tends when the radius goes to zero. This definition may be regarded as a generalization of the above definition for monogenic functions since for such functions it furnishes the same value $R_A$.

Symbolically then we write

$$R' = \lim_{r \to 0} \frac{1}{2\pi i}\int_0 f(z)dz \tag{3.1}$$

for our *definition of the residue of a polygenic function* at the point $z = a$ where the contour is a circle of radius $r$, with center at $a$.

* An arbitrary contour within the circle would serve the purpose just as well.

THEOREM 1. *If $f(z)$ is a polygenic function in a region $A$ and has no singularity at $z=a$ nor in the neighborhood of $z=a$ (and by this we mean that it is possible to construct a circle about the point $a$ which contains no singularity of the function considered, except, possibly, at $z=a$) then*

$$R' = 0. \tag{3.2}$$

This, of course, is the analogue or generalization of Cauchy's first law. We may write

$$\int_0 f(z)dz = \int_0 [f(z) - f(a)]dz + f(a)\int_0 dz.$$

The last integral is evidently zero and the modulus of the preceding will be less than the maximum value of $|f(z)-f(a)|$ on the circle multiplied by $\int_0 |dz|$. Both these factors tend towards zero with $r$ so that $R'=0$.

THEOREM 2. *If $z=a$ is a simple isolated pole of the polygenic function $f(z)$, such that*

$$F(z) = (z - a)f(z)$$

*and $F(z)$ is regular at $a$, and in the neighborhood of $a$, then*

$$R' = F(a). \tag{3.3}$$

This is the generalization of Cauchy's second law for monogenic functions. We may write

$$\int_0 f(z)dz = \int_0 \frac{F(z) - F(a)}{z - a}dz + F(a)\int_0 \frac{dz}{z - a}.$$

The second term in the right member is $2\pi iF(a)$; the first term may be written according to the formula (1.1) as

$$\frac{df}{d\beta}\int_0 dz + \frac{df}{d\alpha}\int_0 \frac{z - a}{\bar{z} - \bar{a}}dz + \int_0 \frac{\eta}{z - a}dz.$$

The first of these integrals vanishes and the other two tend towards zero with the radius of the circle, and the theorem is proved.

We consider now a more difficult problem of finding the residue at a pole of the second type of §2. For convenience we shall take the singularity of the function $f(z)$ at the point $z=0$ and then study the behavior of the function at the origin. We assume then that

$$(az + b\bar{z})f(z) = F(z)$$

where $F(z)$ is regular in the neighborhood of $z=0$. In fact we will treat only the situation where $F(z)$ is unity so that

$$\int_0 f(z)dz = \int_0 \frac{dz}{az + b\bar{z}}. \tag{3.4}$$

We have to consider three different cases,

$$|b| > |a|\,; \quad |b| < |a|\,; \quad |b| = |a|\,.$$

We study together the first two cases.

**Case I,** $|b| > |a|$; **Case II,** $|b| < |a|$. If we change the form of the integral so that

$$\int_0 \frac{dz}{az + b\bar{z}} = \frac{1}{2a}\int_0 \frac{2azdz}{az^2 + br^2}$$

we can easily evaluate it on the circle with a fixed radius $r$. Since in this form the denominator, if $r$ be regarded as fixed, is equal to an analytic function of $z$ and the numerator is its differential, we may apply the theorem in analytic function theory that such an integral is equal to the difference between the number of zeros and the number of poles of the function, in the area bounded by the contour of integration, multiplied by $2\pi i$. In both cases considered, the function $az^2+br^2$ has no poles inside the circle of radius $r$.

In Case I there are two zeros of the function outside the circle and no zeros inside; thus the integral vanishes in this case.

In Case II there are two zeros of the function $az^2+br^2$ inside the circle so that the integral is $2\cdot 2\pi i$. This argument holds for every $r$; we therefore have in the limit

$$R' = \lim_{r\to 0} \frac{1}{2\pi i}\int_0 f(z)dz = \frac{1}{a}.$$

**Case III,** $|b| = |a|$. Here the poles of the function are distributed along a straight line through the origin. The integral in question is improper since the intersections of the line and the circle, radius $r$, make the integrand infinite. In this case $b = ae^{i\phi}$, where $(\phi-\pi)/2$ is the slope angle of the line of poles, and the integral in (3.4) may be transformed as follows:

$$\begin{aligned}
\int_0 \frac{dz}{az + b\bar{z}} &= \frac{1}{a}\int_0 \frac{dz}{z + e^{i\phi}\bar{z}} = \frac{i}{a}\int_0^{2\pi} \frac{e^{i\theta}d\theta}{e^{i\theta} + e^{i(\phi-\theta)}} \\
&= \frac{i}{a}\int_0^{2\pi} \frac{d\theta}{1 + e^{i(\phi-2\theta)}} = \frac{i}{a}\int_0^{2\pi} \frac{d\theta}{1 + \cos(\phi - 2\theta) + i\sin(\phi - 2\theta)} \\
&= \frac{i}{a}\int_0^{2\pi} \frac{[1 + \cos(\phi - 2\theta) + i\sin(\phi - 2\theta)]d\theta}{2 + 2\cos(\phi - 2\theta)} \\
&= \frac{i}{2a}\int_0^{2\pi} d\theta - \frac{1}{2a}\int_0^{2\pi} \tan\left(\theta - \frac{\phi}{2}\right)d\theta.
\end{aligned}$$

Here the first integral evidently becomes $\pi i/a$. We agree to take the *chief* value of the second (improper) integral, letting $\theta$ go from $(\phi-\pi)/2+\epsilon$ to $(\phi+\pi)/2-\epsilon$ and from $(\phi+\pi)/2+\epsilon$ to $(\phi+3\pi)/2-\epsilon$ and then making $\epsilon$ tend toward zero. If we pair off the elements of integration in the first and fourth quadrants and those in the second and third we see that we obtain zero.

If we think of $az+b\bar{z}$ as a function of $\bar{z}$ and define the residue of its reciprocal in the same manner in $\bar{z}$ then $ar^2+b\bar{z}^2$ will be analytic in $\bar{z}$. We thus have a second residue

$$\overline{R} = 1/b$$

in Case I, zero in Case II, and $1/(2b)$ in Case III. If we define the total residue, $R$, as the sum of the residues in $z$ and $\bar{z}$ respectively, we will then have

Case I : $R = 1/b$;

Case II : $R = 1/a$;

Case III : $R = (a + b)/(2ab)$ ·

4. **Areal derivative.** An interesting application of this theory of residues arises in connection with the differentiation of

$$\lim_{r\to 0} \int_0 \frac{f(t)dt}{t-z}$$

under the integral sign. According to Theorem 2 we may write equation (3.3)

$$f(z) = \lim_{r\to 0} \frac{1}{2\pi i} \int_0 \frac{f(t)dt}{t-z} \tag{4.1}$$

where the contour of integration is the circle of radius $r$, center at $z$. This is the generalization of Cauchy's integral representation of a function; in what follows we obtain the generalization for the integral representation of the derivative.

THEOREM 3. *The $\beta$ derivative of a polygenic function may be expressed in the form*

$$\frac{df}{d\beta} = \lim_{r\to 0} \frac{1}{2\pi i} \int_0 \frac{f(t)dt}{(t-z)^2} = \lim_{r\to 0} \frac{1}{2\pi i} \int_0 \frac{d}{dz} \frac{f(t)}{t-z} dt \tag{4.2}$$

*or the $\beta$ derivative may be obtained by taking the $z$ derivative in*

$$\lim_{r\to 0} \int_0 \frac{f(t)dt}{t-z}$$

*under the integral sign.*

If we use formula (1.1) we may write

$$\int_0 \frac{f(t)dt}{(t-z)^2} = \int_0 \frac{f(z)dt}{(t-z)^2} + \left(\frac{df}{d\beta}\right)_z \int_0 \frac{(t-z)dt}{(t-z)^2} + \left(\frac{df}{d\alpha}\right)_z \int_0 \frac{\bar{t}-\bar{z}}{(t-z)^2}dt + \int_0 \frac{\eta dt}{(t-z)^2}.$$

To evaluate the right member we observe that the first integral falls under analytical function theory and that the integral vanishes identically. The last tends towards zero with $r$ since the integrand is finite. In the second and third integrals we make the substitution $re^{i\theta}$ for $t-z$ and obtain

$$\frac{df}{d\beta} i \int_0^{2\pi} d\theta + \frac{df}{d\alpha} i \int_0^{2\pi} \frac{\bar{t}-\bar{z}}{t-z} d\theta.$$

It will be seen that the second integral is identically zero; this leaves $2\pi i(df/d\beta)$ as our limit which proves the theorem.

The $\alpha$ derivative may be found in exactly the same way if we think of the polygenic function $f$ as a function of $\bar{z}$ and write

$$f(\bar{z}) = \lim_{r\to 0} \frac{1}{2\pi i} \int_0 \frac{-f(\bar{t})d\bar{t}}{\bar{t}-\bar{z}}.$$

This equation is the analogue to (3.3) and may be deduced in the same way.

**Conclusion.** In a new field it is frequently difficult to discriminate between the important and the unimportant. However, it would appear that the calculus of residues should play a very important rôle in certain phases of polygenic function theory.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

# ON THE COMMON INDEX DIVISORS OF AN ALGEBRAIC FIELD*

BY

H. T. ENGSTROM

## I. INTRODUCTION

Let $d$ be the discriminant of an algebraic field $K$ generated by a root $\theta$ of an irreducible equation

$$f(x) = x^n + a_1x^{n-1} + \cdots + a_n = 0 \tag{1}$$

whose coefficients are rational integers. We shall call (1) the characteristic equation for $\theta$. If $d_\theta$ is the discriminant of $\theta$ then $d_\theta = k_\theta^2 \cdot d$ where $k_\theta$ is a rational integer, the index of $\theta$. A divisor common to the indices of every integer of the field has been called by Kronecker a "gemeinsamer ausserwesentlicher Discriminantenteiler" of the field. We shall use the term "common index divisor."†

The existence of common index divisors was first established in 1871 by Dedekind‡ who exhibited examples in fields of third and fourth degrees. Dedekind§ further showed that a rational prime $p$ can be a common index divisor of a field $K$ if and only if at least one of the inequalities $r(f) > g(f)$ holds, where $r(f)$ is the number of prime ideal divisors of $p$ of degree $f$, and $g(f)$ is the number of different prime functions (mod $p$) of $f$th degree. Using Kronecker's theory of algebraic numbers, Hensel‖ has given a necessary and sufficient condition on the so-called "index form" for $p$ to be a common index divisor. In 1907, Bauer¶ showed that if $p < n$ there exists a field of $n$th degree in which $p$ is a common index divisor. Von Zylinsky** has

* Presented to the Society, December 27, 1928, and March 30, 1929; received by the editors in June, 1929.

† This terminology corresponds to that of Fricke (*Algebra*, vol. 3, Leipzig, 1928) who uses the term "ständiger Indexteiler."

‡ R. Dedekind, Göttinger Gelehrte Anzeigen, 1871, pp. 1481–1494.

§ R. Dedekind, *Über den Zusammenhang zwischen der Theorie der Ideale und der Theorie der höhere Kongruenzen*, Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen, vol. 23 (1878), pp. 1–23.

‖ K. Hensel, *Arithmetische Untersuchungen über die gemeinsamer ausserwesentlicher Discriminantenteiler einer Gattung*, Journal für Mathematik, vol. 113 (1894), pp. 128–160.

¶ M. Bauer, *Über den ausserwesentlicher Discriminantenteiler algebraischer Körper*, Mathematische Annalen, vol. 64 (1907), p. 573.

** E. von Zylinsky, *Zur Theorie der ausserwesentlicher Discriminantenteiler algebraischer Körper*, Mathematische Annalen, vol. 73 (1913), pp. 273–274.

established the necessity of this condition, i.e., $p$ can be a common index divisor of a field of $n$th degree only if $p<n$.

This paper is concerned with the problem of determining the greatest power, $p^\chi$, of $p$ which is a common index divisor in a given field and, in particular, how $\chi$ is connected with the prime ideal decompositon of $p$. In a recent paper Professor Ore* makes the conjecture that $\chi$ is not in general determined by the prime ideal decomposition of $p$. The author verifies this conjecture by showing that for degrees eight or higher certain types of prime ideal decomposition may occur in fields for which $\chi$ is different. For fields of degree less than eight we show that $\chi$ is determined by the prime ideal decomposition of $p$. Theorems are developed giving this value of $\chi$ for all types of prime ideal decomposition which may occur in such fields. Some types of ideal decomposition determine $\chi$ even in the field of $n$th degree. For some of such cases, including the case where all the ideals are of first degree and first order, $\chi$ is given by the theorems of this paper.

## II. The determination of $\chi$ in some general cases

1. Let $f(x)=0$ be the characteristic equation of an integer $\theta$ of the field $K$ and suppose $d_\theta=p^\delta\cdot q$ where $(p, q)=1$. If

$$f(x)\equiv f_1(x)\cdot f_2(x)\cdots f_r(x) \pmod{p^\alpha}$$

where $\alpha>\delta+1$, and $f_i(x)$, $i=1, 2, 3, 4$, is irreducible (mod $p^\alpha$) and of degree $n_i$, then, by the first theorem of Ore,†

$$[p]=\mathfrak{p}_1^{e_1}\cdot\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_r^{e_r} \tag{2}$$

where $N\mathfrak{p}_i^{e_i}=p^{n_i}$. The field generated by a root $\Theta^{(i)}$ of $f_i(x)=0$ is called the Abbildungskörper of $\mathfrak{p}_i$ and shall be denoted by $K^{(i)}$. Let $p^{\kappa_i}$ denote the highest power of $p$ contained in the index of $\Theta^{(i)}$ in $K^{(i)}$, and furthermore let $p^{\rho_{ij}}$ be the highest power dividing the resultant $R(f_i(x), f_j(x))$. Then the index of $\theta$ is divisible by exactly $p^{\kappa_\theta}$, where

$$\kappa_\theta=\sum_{i>j}\rho_{ij}+\sum_i\kappa_i.\ddagger \tag{3}$$

In order to find $\chi$ for the field $K$ from the prime ideal decomposition (2) it is necessary to determine the values of $\rho_{ij}$ and $\kappa_i$ for an integer $\theta$ of the field

* O. Ore, *Newtonsche Polygone in der Theorie der algebraischen Körper*, Mathematische Annalen, vol. 99 (1928).

† O. Ore, *Über den Zusammenhang zwischen den definierenden Gleichungen und der Idealtheorie in algebraischen Körper*, Mathematische Annalen, vol. 96 (1926).

‡ O. Ore, loc cit., p. 345.

for which $\kappa_\theta$ takes on the least value. We shall first examine the problem for the simplest case.

2. Consider a field $K$ of $n$th degree in which

$$[p] = \mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_n, \tag{4}$$

where $N\mathfrak{p}_i = p$, $i = 1, 2, \cdots, n$. If $f(x) = 0$ is the characteristic equation of an arbitrary integer $\theta$ in $K$, then

$$f(x) \equiv (x + a_1)(x + a_2) \cdots (x + a_n) \pmod{p^\alpha} \tag{5}$$

for $\alpha$ sufficiently great. It is seen that $\kappa_i = 0$, $i = 1, 2, \cdots, n$. Furthermore $R(f_i(x), f_j(x)) = a_i - a_j$. Hence $k_\theta$ is divisible by the same power of $p$ as

$$\prod_{i>j}(a_i - a_j). \tag{6}$$

To determine $\chi$ for the field $K$ we seek a set of rational integers $a_1, a_2, \cdots, a_n$ which correspond to an integer of the field by (5) and such that (6) contains the least possible power of $p$. We shall make use of the following theorem:

THEOREM 1. *Let $\mathfrak{p}_1, \mathfrak{p}_2, \cdots, \mathfrak{p}_s$ be any distinct prime ideal divisors of $p$ in $K$ and let $\phi_1(x), \phi_2(x), \cdots, \phi_s(x)$ be a set of prime functions, not necessarily distinct, of degrees $m_1, m_2, \cdots, m_s$ where $m_i$ divides the degree $f_i$ of $\mathfrak{p}_i$. Then there exists a primitive integer $\theta$ in $K$ such that $\phi_i(\theta)$ contains exactly $\mathfrak{p}_i^{h_i}$, $i = 1, 2, \cdots, s$, where $h_1, h_2, \cdots, h_s$ is an arbitrary set of positive rational integers.*

For an arbitrary prime ideal $\mathfrak{p}$ of degree $f$ dividing $p$, every integer $\omega$ of $K$ satisfies the congruence

$$\omega^{p^f} - \omega \equiv 0 \pmod{\mathfrak{p}}$$

which has, therefore, as many roots as its degree $p^f$. On the other hand, the polynomial $x^{p^f} - x$ is congruent (mod $p$) to the product of all prime functions (mod $p$) whose degrees are divisors of $f$. It follows that

$$\phi_1(x) \equiv (x - \omega_1)(x - \omega_2) \cdots (x - \omega_{m_1}) \pmod{\mathfrak{p}_1}$$

where $\omega_1, \omega_2, \cdots, \omega_{m_1}$ are integers in $K$. By a generalization of a theorem of Schönemann* we have

$$\phi_1(x) \equiv (x - \mu_1)(x - \mu_2) \cdots (x - \mu_{m_1}) \pmod{\mathfrak{p}_1^{h_1}}$$

where $\mu_1, \mu_2, \cdots, \mu_{m_1}$ are also integers in $K$. Hence there exists an integer

* Th. Schönemann, *Von denjenigen Moduln, welche Potenzen von Primzahlen sind*, Journal für Mathematik, vol. 32 (1846), pp. 93–95.

$\theta_1$ in $K$ such that $\phi_1(\theta_1) \equiv 0 \pmod{\mathfrak{p}_1^{h_1}}$. We may suppose $\theta_1$ chosen so that $\phi_1(\theta_1)$ is divisible by exactly $\mathfrak{p}_1^{h_1}$, for suppose $\phi_1(\theta_1)$ were divisible by $\mathfrak{p}_1^{\beta}$, $\beta > h_1$. Then if $\pi$ is an integer in $K$ containing exactly $\mathfrak{p}_1^1$ and we set $\theta_1' = \theta_1 + \pi^{h_1}$, we have

$$\phi_1(\theta_1') = \phi_1(\theta_1 + \pi^{h_1}) = \phi_1(\theta_1) + \pi^{h_1} \cdot \phi_1'(\theta_1) + \cdots + \frac{\pi^{mh_1}}{m!}\phi_1^{(m)}(\theta_1).$$

Since $\phi_1'(\theta_1) \not\equiv 0 \pmod{\mathfrak{p}_1}$, it follows that $\phi_1(\theta_1')$ is divisible by exactly $\mathfrak{p}_1^{h_1}$.

Similarly we may determine integers $\theta_2, \theta_3, \cdots, \theta_s$ so that $\phi_i(\theta_i)$ is divisible by exactly $\mathfrak{p}_i^{h_i}$, $i = 1, 2, \cdots, s$, and hence if $\theta$ is chosen so that

$$\theta \equiv \theta_i \pmod{\mathfrak{p}_i^{h_i+1}} \qquad (i = 1, 2, \cdots, s),$$

it follows that $\phi_i(\theta)$ is divisible by exactly $\mathfrak{p}_i^{h_i}$, $i = 1, 2, \cdots, s$. Furthermore $\theta$ may be supposed primitive, for if it were not primitive let

$$\theta' = \theta + p^{h+1} \cdot r \cdot \Theta \qquad (h > h_i, i = 1, 2, \cdots, s),$$

where $\Theta$ is a primitive integer and $r$ is a rational integer. It is seen that $r$ can be chosen so that $\theta'$ is primitive while $\phi_i(\theta')$ contains exactly $\mathfrak{p}_i^{h_i}$, $i = 1, 2, \cdots, s$. Hence Theorem 1 is proved. From Theorem 1 we have the following Theorem:

THEOREM 2. *If $a_1, a_2, \cdots, a_n$ is an arbitrary set of $n$ rational integers, then there exists an integer $\theta$ in $K$ for which $f(x)$ has the decomposition* (5) *for some $\alpha$.*

For by Theorem 1 we may choose $\theta$ so that $\theta + a_i \equiv 0 \pmod{\mathfrak{p}_i^{\beta}}$, $i = 1, 2, \cdots, n$. Suppose

$$f(x) \equiv (x + b_1)(x + b_2) \cdots (x + b_n) \pmod{p^{\alpha}}.$$

Then $\theta + b_i \equiv 0 \pmod{\mathfrak{p}_i^{\alpha - \rho'}}$ where $\alpha$ may be chosen so large that $\alpha - \rho' > \beta$. Hence $\theta + a_i \equiv \theta + b_i \pmod{\mathfrak{p}_i^{\beta}}$ and $a_i \equiv b_i \pmod{p^{\beta}}$ and Theorem 2 follows. If the $a$'s are all distinct and $\beta$ is chosen so that they are distinct (mod $p^{\beta}$) then

$$f(x) \equiv (x + a_1)(x + a_2) \cdots (x + a_n) \pmod{p^{\beta}}$$

where $\beta > \delta$.

To determine $\theta$ so that (6) contains the least power of $p$ we make use of the following theorem due to Hensel.*

*The discriminant $\prod_{i \neq k} (a_i - a_k)$ of $n$ arbitrary rational integers is divisible by the discriminant $D_n = \prod_{i \neq j} (i - j)$ of the first $n$ positive rational integers.*

Hence, for a field in which $p$ has the decomposition (4), $D_n$ contains exactly $p^{\chi}$. The following theorem is obtained by simple calculation.

* K. Hensel, *Über den grössten gemeinsamen Teiler aller Zahlen welche durch ein ganze Funktion von n Veränderlichen darstellbar sind*, Journal für Mathematik, vol. 116 (1896), p. 352.

THEOREM 3. *If $K$ is an algebraic field of nth degree in which*

$$[p] = \mathfrak{p}_1 \cdot \mathfrak{p}_2 \cdots \mathfrak{p}_n, \quad N\mathfrak{p}_i = p,$$

*then*

$$\chi = \sum_{i=1} s_i \left\{ n - p^i \left( \frac{s_i + 1}{2} \right) \right\}, \quad s_i = \left[ \frac{n}{p^i} \right].$$

3. We shall extend this result to a more general type of prime ideal decomposition. Suppose $\mathfrak{p}_i$ is a prime ideal of order $e_i$ and degree $f_i$ dividing $p$, and $\phi(x)$ is a prime function (mod $p$) of degree $f_i$. We may determine a primitive integer $\theta$ in $K$ such that $\phi(\theta)$ is divisible by exactly $\mathfrak{p}_i^1$. If $f_i(x)$ is the factor (mod $p^a$) corresponding to $\mathfrak{p}_i$ of the characteristic equation for $\theta$, then the Newton polygon $(p, \phi(x))$ of $f_i(x)$ is a straight line with slope $1/e_i$.* Since $f_i(x)$ must be of degree $e_i f_i$ we have

$$f_i(x) = \phi(x)^{e_i} + p \cdot M(x), \tag{7}$$

where $M(x) \not\equiv 0$ (modd $p, \phi(x)$). After Ore, we call the form (7) the "normal form." It follows by the theorem of Dedekind that if the factor corresponding to an ideal $\mathfrak{p}_i$ is normal then $\kappa_i = 0$. If the prime ideal decomposition of $p$ is such that $g(f_i) < r(f_i)$ for $f_i \neq 1$, it follows that we may choose $\theta$ so that $\kappa_i = 0$ for $f_i \neq 1$. Furthermore, if $f_i \neq 1$, $R(f_i(x), f_j(x)) \not\equiv 0$ (mod $p$), i.e. $\rho_{ij} = 0$. Hence we have the following theorem:

THEOREM 4. *Let $K$ be an algebraic field of nth degree in which*

$$[p] = \mathfrak{p}_1^{e_1} \cdot \mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_s^{e_s}, \quad N\mathfrak{p}_i = p^{f_i}.$$

*If $g(f_i) < r(f_i)$ for $f_i \neq 1$, and $e_i = 1$ for $f_i = 1$, then*

$$\chi = \sum_i s_i \left\{ r - p^i \left( \frac{s_i + 1}{2} \right) \right\}, \quad s_i = \left[ \frac{r}{p^i} \right],$$

*where $r$ is the number of first degree prime ideals dividing $p$.*

4. If the degree of the field is less than eight, Theorem 4 can be extended to the case where some of the first degree ideals are of higher order. We shall say that a prime ideal $\mathfrak{p}$ of $f$th degree dividing $p$ is associated with a prime function $\phi(x)$ (mod $p$), whose degree is a divisor of $f$, for the integer $\theta$ if $\mathfrak{p}$ divides $\phi(\theta)$. It follows from Theorem 1 that an integer $\theta$ can be found in any field with any given association of the prime ideal divisors of $p$. We shall

* O. Ore, *Newtonsche Polygone in der Theorie der algebraischen Körper*, Mathematische Annalen, vol. 99 (1928), p. 100.

use the notation $\mathfrak{p}^a \mid \phi(\theta)$ to indicate that $\phi(\theta)$ is divisible by exactly the $a$th power of $\mathfrak{p}$. Let

$$\mathfrak{p}_1, \mathfrak{p}_2, \cdots, \mathfrak{p}_t \tag{8}$$

be the prime ideals dividing $p$ for which $f_i = 1$ and $e_i > 1$, and let

$$\mathfrak{q}_1, \mathfrak{q}_2, \cdots, \mathfrak{q}_r \tag{9}$$

be those of first order and first degree. Suppose further that $g(f_i) \geqq r(f_i)$ for $f_i \neq 1$. If the ideals (8) are each associated with a distinct prime function of first degree for an integer $\theta$ whose index contains exactly $p^\chi$, then $\chi$ may be determined from the prime ideal decomposition. We shall show that this association can be assumed for $\theta$ in a field of degree less than eight if $t \leqq p$.

Suppose that $t \leqq p$ and that at least two of the ideals (8), say $\mathfrak{p}_1$ and $\mathfrak{p}_2$, must necessarily be associated with a single prime function, say $x$, for an integer $\theta$ whose index is divisible by exactly $p^\chi$. Then there must exist at least $p-t+1$ prime functions, say

$$x+1, x+2, \cdots, x+p-t+1,$$

with which none of the ideals (8) are associated for $\theta$. Suppose that $n_i$ of the ideals (9) divide $\theta+i$, $i=1, 2, \cdots, p+t-1$. If $f_i(x)$ is the factor (mod $p^a$) of $f(x)$ corresponding to $\mathfrak{p}_i$, we have, from the Newton polygon,

$$f_i(x) = x^{e_i} + pM_i(x), \quad i = 1, 2.$$

Hence $\rho_{12} \geqq 2$.

Consider the value of $\kappa$ for another integer $\theta_1$ in $K$ for which $\mathfrak{p}_1^1 \mid \theta_1 + 1$ while the other associations remain the same as for $\theta$. Let $f_1^{(1)}(x)$ be the factor (mod $p^a$) of the characteristic equation of $\theta_1$ which corresponds to $\mathfrak{p}_1$. Then

$$f_1^{(1)}(x) = (x+1)^{e_1} + p \cdot B, \quad B \not\equiv 0 \pmod{p}.$$

Hence for $\theta_1$ we have $\kappa_1 + \sum \rho_{1j} = n_1$. It follows from our assumption that the value of $\kappa$ for $\theta_1$ must exceed $\chi$, i.e.

$$n_1 > 2. \tag{10}$$

Similarly, $n_i > 2$ for $i = 1, 2, \cdots, p-t+1$. Hence if $n$ is the degree of the field, we have

$$n \geqq 3(p-t+1) + 2t. \tag{11}$$

It follows that $n \geqq 7$, i.e., if $t \leqq p$ and $n \leqq 7$ the index of some integer $\theta$ for which the ideals (8) are associated with distinct prime functions is divisible by exactly $p^\chi$.

This result is true for $n=7$. To establish this suppose first that $t\geqq 3$. It follows from (11) that $n\geqq 9$. In the second place if $t=2$ and $p>2$ from (11) we have $n\geqq 10$. We have yet to consider the case where $t=2$ and $p=2$. In the field of seventh degree this becomes the case where

$$[2] = \mathfrak{p}_1^2\cdot\mathfrak{p}_2^2\cdot\mathfrak{p}_3\cdot\mathfrak{p}_4\cdot\mathfrak{p}_5, \quad N\mathfrak{p}_i = 2. \tag{12}$$

Let us assume as above that $\chi$ must necessarily occur for an integer $\theta$ in $K$ for which both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are associated with a single prime function, say $x$. From (10) it follows that the other ideals $\mathfrak{p}_3$, $\mathfrak{p}_4$, $\mathfrak{p}_5$ must be associated with $x+1$ for $\theta$. Suppose $\mathfrak{p}_1^a\cdot\mathfrak{p}_2^b\,|\,\theta$. Then

$$f(x) \equiv f_1(x)f_2(x)\cdots f_5(x) \pmod{2^a},$$

where

$$\begin{aligned} f_1(x) &= x^2 + 2k_1x + 2^am_1, & (m_1,2) &= 1,\\ f_2(x) &= x^2 + 2k_2x + 2^bm_2, & (m_2,2) &= 1. \end{aligned} \tag{13}$$

We shall make use of the following lemma:

LEMMA 1. *If $f_1(x)$ and $f_2(x)$ have the form* (13), *then* $\min(\kappa_1+\kappa_2+\rho_{12})=3$.

For we have

$$R(f_1(x),\ f_2(x)) = -(2^am_1 - 2^bm_2)^2 + 2(k_2-k_1)(k_12^bm_2 - k_22^am_1). \tag{14}$$

If $a\geqq 2$ and $b\geqq 2$, it follows that $\rho_{12}\geqq 3$. If $a=1$ and $b>1$ we have from (14) that $\rho_{12}=2$. Furthermore $\Theta^{(2)}/2$ is an integer in the Abbildungskörper $K^{(2)}$ since it satisfies the equation

$$x^2 + k_2x + 2^{b-2}m_2 = 0.$$

Hence $\kappa_2\geqq 1$, i.e., $\kappa_1+\kappa_2+\rho_{12}\geqq 3$. The similar result follows if $a>1$, $b=1$. If $a=1$ and $b=1$ we have from (14) that $\rho_{12}\geqq 3$. We have thus shown that $\min(\kappa_1+\kappa_2+\rho_{12})\geqq 3$.

It remains to show that there exists an integer $\theta$ in the field for which $\kappa_1+\kappa_2+\rho_{12}=3$. By choosing $a=1$ and $b=3$ we have $\rho_{12}=2$ and furthermore $\kappa_1=0$ since $f_1(x)$ is in the normal form. It is also seen that $\Theta^{(2)}/2$ satisfies an equation which has the normal form and hence $\kappa_2=1$. For this choice of $\theta$, $\kappa_1+\kappa_2+\rho_{12}=3$ and the lemma follows.

Returning to our argument we shall show that the assumption that $\mathfrak{p}_1$ and $\mathfrak{p}_2$ in (12) must be associated with the same prime function for the $\theta$ whose index contains exactly $2^x$ leads to a contradiction. For if an integer $\theta_1$ is chosen so that $\mathfrak{p}_2^1\,|\,\theta_1+1$ while the other associations are the same as for $\theta$, we would have

$$\kappa_2 + \sum_i \rho_{2i} = 3,$$

the other terms in (3) remaining the same as for $\theta$. We have thus shown that in a field of degree less than eight, $t \leqq p$, the index of some integer for which each of the ideals (8) is associated with a distinct prime function contains exactly $p^{\chi}$.

It follows that the required integer is such that the ideals (9) are associated with prime functions as in Theorem 4 while each of the ideals (8) is associated with a distinct prime function, first exhausting the prime functions with which the smaller number of ideals (9) is associated. The following theorem is obtained by simple calculation.

THEOREM 5. *If $K$ is an algebraic field of degree $n<8$ in which*

$$[p] = \mathfrak{p}_1^{e_1} \cdot \mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_s^{e_s}, \quad N\mathfrak{p}_i = p^{f_i},$$

*where $g(f_i) > r(f_i)$ for $f_i \neq 1$, and $e_i = 1$ for $f_i = 1$ except for $t$ ideals, $t \leqq p$, then*

$$\chi = t(s_1 + 1) - \min(t, p - r + ps_1) + \sum_i s_i \left\{ r - p^i \left( \frac{s_i + 1}{2} \right) \right\}, \quad s_i = \left[ \frac{r}{p^i} \right],$$

*where $r$ is the number of first degree, first order prime ideals dividing $p$.*

If $t = 1$ it follows immediately that the above result is valid for a field of $n$th degree and we have the following corollary:

COROLLARY. *If $K$ is a field of $n$th degree in which $t = 1$ and the other conditions of Theorem 5 are satisfied, then*

$$\chi = s_1 + \sum_i s_i \left\{ r - p \left( \frac{s_i + 1}{2} \right) \right\}. \tag{15}$$

5. We shall now consider certain types of ideal decomposition of the prime 2 for which $\chi$ may be determined. Let $K$ be a field of $n$th degree in which

$$[2] = \mathfrak{p}_1^a \cdot \mathfrak{p}_2^b \cdot \mathfrak{p}_3^2, \quad N\mathfrak{p}_i = 2,$$

and $a \geqq b \geqq 2$. Choose an integer $\theta$ so that $\mathfrak{p}_1^1 \mathfrak{p}_3^c \mid \theta$ and $\mathfrak{p}_2 \mid \theta + 1$. Then the characteristic equation of $\theta$ has the decomposition

$$f(x) \equiv f_1(x) f_2(x) f_3(x) \pmod{2^a},$$

where

$$f_1(x) = x^a + 2k_1 x^{a-1} + \cdots + 2m_1, \quad (m_1, 2) = 1,$$
$$f_2(x) = (x+1)^b + 2l_1(x+1)^{b-1} + \cdots + 2m_2, \quad (m_2, 2) = 1,$$
$$f_3(x) = x^2 + 2n_1 x + 2^c m_3, \quad (m_3, 2) = 1.$$

If $a=2$, it follows from Lemma 1 that min $(\rho_{13}+\kappa_1+\kappa_3)=3$ for integers of $K$ for which $\mathfrak{p}_1$ and $\mathfrak{p}_3$ are associated with $x$. For the $\theta$ chosen above we have $\kappa_1=\kappa_2=\rho_{12}=0$, and furthermore, as in Lemma 1, if we choose $c=3$ it follows that $\kappa_3=1$ and $\rho_{13}=2$. Hence, if $a=2$, we have $\chi=3$.

Consider the case where $a>2$. Since $R_{ij}=N^{(i)}(f_j(\Theta^{(i)}))$, where $N^{(i)}$ indicates the norm taken in the Abbildungskörper $K^{(i)}$, it follows that $\rho_{ij}\geqq 2$ for the two ideals $\mathfrak{p}_i$ and $\mathfrak{p}_j$ which are associated with the same prime function. Hence we have min $(\rho_{13}+\kappa_1+\kappa_3)\geqq 2$. But if $c=1$, we have

$$R_{13} = N^{(3)}(\Theta^{(3)a} + 2k_1\Theta^{(3)a-1} + \cdots + 2m_1) \tag{16}$$

where

$$\Theta^{(3)2} + 2n_1\Theta^{(3)} + 2m_3 = 0. \tag{17}$$

From (17) it follows that $\mathfrak{p}_3^1 \mid \Theta^{(3)}$. Hence from (16) we have $\mathfrak{p}_3^4 \mid R_{13}$, i.e. $\rho_{13}=2$. Furthermore $\kappa_1=\kappa_2=\kappa_3=\rho_{12}=\rho_{23}=0$. Hence for this choice of $\theta$ we have $\kappa=2$ and we conclude that $\chi=2$ for $a>2$.

This result may be extended to the ideal decomposition

$$[2] = \mathfrak{p}_1^a \cdot \mathfrak{p}_2^b \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4, \quad N\mathfrak{p}_i = 2, \quad a \geqq b \geqq 2.$$

If $\theta$ is chosen so that $\mathfrak{p}_1^1\mathfrak{p}_3^3 \mid \theta$ and $\mathfrak{p}_2^1\mathfrak{p}_4^1 \mid \theta+1$ it is seen as above that $2^x \mid k\theta$, i.e., $\chi=3$ or 4 according as $a>2$ or $a=2$. Similarly for the field in which

$$[2] = \mathfrak{p}_1^a \cdot \mathfrak{p}_2^b \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4 \cdot \mathfrak{p}_5, \quad N\mathfrak{p}_i = 2, \quad a \geqq b \geqq 2,$$

we choose $\theta$ so that $\mathfrak{p}_1^1\mathfrak{p}_3^3 \mid \theta$ and $\mathfrak{p}_2^1 \cdot \mathfrak{p}_4^1 \cdot \mathfrak{p}_5^2 \mid \theta+1$ and obtain $\chi=5$ or 6 according as $a>2$ or $a=2$. These results may be combined in the following theorem:

THEOREM 6. *If $K$ is a field of nth degree in which*

$$[2] = \mathfrak{p}_1^a \cdot \mathfrak{p}_2^b \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4^c \cdot \mathfrak{p}_5^d, \quad N\mathfrak{p}_i = 2,$$

*where $a\geqq b\geqq 2>c\geqq d$, then $\chi=2+c+2d$ or $3+c+2d$ according as $a>2$ or $a=2$.*

This theorem is also true in fields in which [2] contains prime ideals of degree greater than one, provided that $r(f)\leqq g(f)$ for $f\neq 1$.

6. Consider a field in which

$$[2] = \mathfrak{p}_1^{(2)a} \cdot \mathfrak{p}_2^{(2)} \cdot \mathfrak{p}_3^b \cdot \mathfrak{p}_4 \cdot \mathfrak{p}_5 \cdots \mathfrak{p}_s$$

where $N\mathfrak{p}_1^{(2)} = N\mathfrak{p}_2^{(2)} = 2^2$; $N\mathfrak{p}_i = 2$ for $i > 2$. It is easily shown that $\phi(x) = x^2 + x + 1$ is the only prime function (mod 2) of second degree. Suppose an integer $\theta$ is chosen so that $\mathfrak{p}_1^{(2)^1} \cdot \mathfrak{p}_2^{(2)^2} | \phi(\theta)$. Then if $f_1(x)$ and $f_2(x)$ are the factors (mod $2^a$) of the characteristic equation of $\theta$ which correspond to $\mathfrak{p}_1^{(2)}$ and $\mathfrak{p}_2^{(2)}$, we have

$$f_1(x) = \phi(x)^a + Q_1(x)2^{\beta_1}\phi(x)^{a-1} + \cdots + 2Q_a(x),$$
$$f_2(x) = \phi(x) + 2^2 R(x),$$

where $Q_i(x) \not\equiv 0 \pmod{2, \phi(x)}$ and $R(x) \not\equiv 0 \pmod{2, \phi(x)}$. It follows that $\kappa_1 = \kappa_2 = 0$ and $\rho_{12} = 2$. If $\mathfrak{p}_1^{(2)}$ and $\mathfrak{p}_2^{(2)}$ are not associated with first degree prime functions it follows from the corollary to Theorem 5 that $\kappa$ is least for the integer $\theta$ for which the ideals $\mathfrak{p}_4, \mathfrak{p}_5, \cdots, \mathfrak{p}_s$ are associated as in that theorem. Hence the least value of $\kappa$ for an integer so chosen is $2+T$ where $T$ is the expression (15). If, on the other hand, $\theta$ is chosen so that one of the ideals $\mathfrak{p}_1^{(2)}$, $\mathfrak{p}_2^{(2)}$, say $\mathfrak{p}_1^{(2)}$, is associated with a first degree prime function, say $x$, then

$$f_1(x) = x^{2a} + 2q_1 x^{2a-1} + \cdots + 2^2 q_{2a},$$

which is not in the normal form and hence $\kappa_1 \geqq 1$. But if $s \geqq 4$ the association of Theorem 5 has at least one of the ideals $\mathfrak{p}_1, \mathfrak{p}_2, \cdots, \mathfrak{p}_s$ associated with $x$ and hence $\sum_j \rho_{ij} \geqq 2$. Furthermore, any significant change of association of the ideals $\mathfrak{p}_3, \mathfrak{p}_4, \cdots, \mathfrak{p}_s$ would increase $\kappa$ by at least one and, since $\kappa_1 \geqq 1$, we would have $\kappa \geqq 2+T$. It follows that if $s \geqq 4$, $\chi = 2+T$ where $T$ is the expression (15).

If $s < 4$, let us determine $\theta$ so that $\mathfrak{p}_1^1 | \theta^2+\theta+1$, $\mathfrak{p}_2^1 | \theta$, and $\mathfrak{p}_3^1 | \theta+1$. Then

$$f_1(x) = \phi(x)^a + 2^{\beta_1} Q_1(x)\phi(x)^{a-1} + \cdots + 2Q_a(x),\ Q_i(x) \not\equiv 0 \pmod{2, \phi(x)};$$

(18) $$f_2(x) = x^2 + 2kx + 4m, \quad (m, 2) = 1;$$

$$f_3(x) = (x+1)^b + 2M(x), \quad M(x) \not\equiv 0 \pmod{2, x+1}.$$

We may further suppose that $\theta$ satisfies the congruence*

$$\theta \equiv 2\omega \pmod{\mathfrak{p}_2^{(2)^a}} \tag{19}$$

where

$$\omega^2 + \omega + 1 \equiv 0 \pmod{\mathfrak{p}_2^{(2)^a}}. \tag{20}$$

* O. Ore, *Über den Zusammenhang zwischen den definierenden Gleichungen und der Idealtheorie in algebraischen Körpern*, Mathematische Annalen, vol. 97 (1927), p. 576.

From (19) and (20) we have $\theta^2+2\theta+4\equiv 0 \pmod{\mathfrak{p}_2^a}$. But from (18) $\theta^2+2k\theta+4m\equiv 0 \pmod{\mathfrak{p}_2^{(2)^{\alpha-\rho}}}$ where we may assume $\alpha-\rho>2$. It follows that $2(k-1)\theta\equiv 4(1-m) \pmod{\mathfrak{p}_2^{\alpha-\rho}}$, and hence $k$ must be odd. The integer $\Theta^{(2)}/2$ in the Abbildungskörper $K^{(2)}$ satisfies the equation $x^2+kx+m=0$, which, since $k$ is odd, has the normal form $x^2+x+1+2M(x)=0$. Hence $\kappa_2=1$ and, since $\kappa_1=\kappa_3=\rho_{1j}=0$, $\chi=1$. We thus have the following theorem:

THEOREM 7. *In a field of nth degree in which*

$$[2] = \mathfrak{p}_1^{(2)^a}\cdot\mathfrak{p}_2^{(2)}\cdot\mathfrak{p}_3^b\cdot\mathfrak{p}_4\cdots\mathfrak{p}_s,$$

$$N\mathfrak{p}_1 = N\mathfrak{p}_2 = 2^2; \quad N\mathfrak{p}_i = 2,\ i>2,$$

*we have*

$$\chi = 2 + s_1 + \sum_{i=1} s_1\left\{r - 2^i\left(\frac{s_i+1}{2}\right)\right\} \text{ or } \chi = 1$$

*according as* $s\geqq 4$ *or* $2\leqq s<4$.

7. As a last case consider a field in which

$$[2] = \mathfrak{p}_1^{(2)}\cdot\mathfrak{p}_2^{(2)}\cdot\mathfrak{p}_3^{(2)}\cdot\mathfrak{p}_4^a,$$

$$N\mathfrak{p}_i = 2^2,\ i = 1,2,3;\ N\mathfrak{p}_4 = 2.$$

If $a=0$ choose $\theta$ so that $\mathfrak{p}_1^{(2)^1}|\theta^2+\theta+1$, $\mathfrak{p}_2^{(2)^1}|\theta$ and $\mathfrak{p}_3^{(2)^1}|\theta+1$. As above, $\rho_{ij}=0$ and $\kappa_2=\kappa_3=1$. It is seen that the index of $\theta$ contains exactly $2^x$ since any other association leads to a larger $\kappa$, i.e. $\chi=2$. Similarly, if $a\neq 0$, by choosing $\theta$ so that $\mathfrak{p}_1\cdot\mathfrak{p}_2^2|\theta^2+\theta+1$, $\mathfrak{p}_3|\theta$ and $\mathfrak{p}_4|\theta+1$, we see that $\chi=3$. Hence we have the following theorem:

THEOREM 8. *If K is a field of nth degree in which*

$$[2] = \mathfrak{p}_1^{(2)}\cdot\mathfrak{p}_2^{(2)}\cdot\mathfrak{p}_3^{(2)}\cdot\mathfrak{p}_4^a,$$

$$N\mathfrak{p}_i = 2^2,\ i = 1,2,3;\ N\mathfrak{p}_4 = 2;$$

*then* $\chi=2$ *or* 3 *according as* $a=0$ *or* $a\neq 0$.

8. The theorem of von Zylinsky and the criterion of Dedekind are sufficient for determining which rational primes are common index divisors in a field from the prime ideal decompositions. The theorems of this paper are sufficient for calculating the powers of these primes contained in the greatest common index divisor of any field of degree less than eight. The following table gives the results of these calculations.

POWERS OF RATIONAL PRIMES CONTAINED IN THE GREATEST COMMON INDEX DIVISOR OF AN ALGEBRAIC FIELD

$[p] = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_r^{e_r}$, $N\mathfrak{p}_i = p^{f_i}$

| Degree of Field | Prime Ideal Decomposition $[f_1, f_2, f_3, \cdots]$ | $[e_1, e_2, e_3, \cdots]$ | Prime 2 | 3 | 5 |
|---|---|---|---|---|---|
| | | | x | x | x |
| 3 | [1 1 1] | [1 1 1] | 1 | | |
| 4 | [1 1 1 1] | [1 1 1 1] | 2 | 1 | |
| | [1 1 1] | [2 1 1] | 1 | | |
| | [2 2] | [1 1] | 1 | | |
| 5 | [1 1 1 1 1] | [1 1 1 1 1] | 5 | 2 | |
| | [1 1 1 1] | [2 1 1 1] | 2 | 1 | |
| | [1 1 1] | [2 2 1] | 1 | | |
| | [1 1 1] | [3 1 1] | 1 | | |
| | [2 1 1 1] | [1 1 1 1] | 1 | | |
| | [2 2 1] | [1 1 1] | 1 | | |
| 6 | [1 1 1 1 1 1] | [1 1 1 1 1 1] | 8 | 3 | 1 |
| | [1 1 1 1 1] | [2 1 1 1 1] | 4 | 2 | |
| | [1 1 1] | [2 2 2] | 3 | | |
| | [1 1 1 1] | [2 2 1 1] | 2 | 1 | |
| | [1 1 1 1] | [3 1 1 1] | 2 | 1 | |
| | [2 1 1 1 1] | [1 1 1 1 1] | 2 | 1 | |
| | [2 2 1 1] | [1 1 1 1] | 2 | | |
| | [2 2 2] | [1 1 1] | 2 | | |
| | [1 1 1] | [3 2 1] | 1 | | |
| | [2 1 1 1] | [1 2 1 1] | 1 | | |
| | [2 2 1] | [1 1 2] | 1 | | |
| | [2 2] | [2 1] | 1 | | |
| | [3 1 1 1] | [1 1 1 1] | 1 | | |

| Degree of Field | Prime Ideal Decomposition $[f_1, f_2, f_3, \cdots]$ | $[e_1, e_2, e_3, \cdots]$ | Prime 2 $\chi$ | 3 $\chi$ | 5 $\chi$ |
|---|---|---|---|---|---|
| 7 | [1 1 1 1 1 1 1] | [1 1 1 1 1 1 1] | 12 | 4 | 2 |
| | [1 1 1 1 1 1] | [2 1 1 1 1 1] | 7 | 3 | 1 |
| | [2 1 1 1 1 1] | [1 1 1 1 1 1] | 5 | 2 | |
| | [1 1 1 1 1] | [2 2 1 1 1] | 4 | 2 | |
| | [1 1 1 1 1] | [3 1 1 1 1] | 4 | 2 | |
| | [1 1 1 1] | [2 2 2 1] | 4 | 1 | |
| | [2 2 1 1 1] | [1 1 1 1 1] | 3 | | |
| | [2 2 2 1] | [1 1 1 1] | 3 | | |
| | [1 1 1 1] | [3 2 1 1] | 2 | 1 | |
| | [1 1 1] | [3 2 2] | 2 | | |
| | [1 1 1 1] | [4 1 1 1] | 2 | 1 | |
| | [2 1 1 1 1] | [1 2 1 1 1] | 2 | 1 | |
| | [2 2 1 1] | [1 1 2 1] | 2 | | |
| | [3 1 1 1 1] | [1 1 1 1 1] | 2 | 1 | |
| | [1 1 1] | [3 3 1] | 1 | | |
| | [1 1 1] | [4 2 1] | 1 | | |
| | [2 1 1 1] | [1 2 2 1] | 1 | | |
| | [2 1 1 1] | [1 3 1 1] | 1 | | |
| | [2 1 1 1] | [2 1 1 1] | 1 | | |
| | [2 2 1] | [1 2 1] | 1 | | |
| | [2 2 1] | [1 1 3] | 1 | | |
| | [3 1 1 1] | [1 2 1 1] | 1 | | |
| | [3 2 2] | [1 1 1] | 1 | | |
| | [4 1 1 1] | [1 1 1 1] | 1 | | |

## III. An example showing that $\chi$ is not determined by the prime ideal decomposition in the general case

Consider a field $K$ of eighth degree in which

$$[3] = \mathfrak{p}_1^2 \cdot \mathfrak{p}_2^2 \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4^2, \quad N\mathfrak{p}_i = 3. \tag{21}$$

Let $\theta$ be an integer in $K$ chosen so that

$$\mathfrak{p}_1^1 \mathfrak{p}_2^1 \mid \theta, \quad \mathfrak{p}_3^1 \mid \theta + 1, \quad \mathfrak{p}_4^1 \mid \theta + 2. \tag{22}$$

If $f(x) = 0$ is the characteristic equation for $\theta$, then

$$f(x) \equiv f_1(x) f_2(x) f_3(x) f_4(x) \pmod{3^a},$$

where, from the Newton polygons,

$$
\begin{aligned}
f_1(x) &= x^2 + 3k_1x + 3m_1, \\
f_2(x) &= x^2 + 3k_2x + 3m_2, \\
f_3(x) &= (x+1)^2 + 3k_3(x+1) + 3m_3, \\
f_4(x) &= (x+2)^2 + 3k_4(x+1) + 3m_4,
\end{aligned} \tag{23}
$$

and $(m_i, 3) = 1$. Then $\rho_{i3} = \rho_{i4} = \kappa_i = 0$, $i = 1, 2, 3, 4$. To calculate $\kappa_0$ we need only calculate $\rho_{12}$. We have

$$R(f_1(x), f_2(x)) \equiv -9(m_2 - m_1)^2 \pmod{3^3}, \tag{24}$$

and therefore $\rho_{12} = 2$ or $\rho_{12} > 2$ according as $m_1 \not\equiv m_2 \pmod 3$ or $m_1 \equiv m_2 \pmod 3$.

We shall show that $m_1 \pmod 3$ is independent of the choice of $\theta$ satisfying (22). Since $\theta$ is a prime with respect to $\mathfrak{p}_1$, any integer $\theta'$ of $K$ satisfies a congruence

$$\theta' \equiv a + b\theta \pmod{\mathfrak{p}_1^{3^\alpha}} \tag{25}$$

where $a$ and $b$ are rational integers. If $\theta'$ satisfies (22) we must have $a \equiv 0 \pmod{\mathfrak{p}_1}$ and $b \not\equiv 0 \pmod{\mathfrak{p}_1}$, i.e., $a \equiv 0 \pmod 3$ and $b \not\equiv 0 \pmod 3$. Let $f_1^{(1)}(x)$ be the factor corresponding to $\mathfrak{p}_1$ in the decomposition (mod 3) of the characteristic equation of $\theta'$. Then $f_1^{(1)}(x) = x^2 + 3k_1'x + 3m_1'$. Hence

$$\theta'^2 + 3k_1'\theta' + 3m_1' \equiv 0 \pmod{\mathfrak{p}_1^\alpha}. \tag{26}$$

Substituting (25) in (26) and using the relations $a \equiv 0 \pmod 3$ and $\theta^2 + 3k_1\theta + 3m_1 \equiv 0 \pmod{\mathfrak{p}_1^{\alpha-\rho}}$, we have

$$3m_1b^2 - 3m_1' \equiv 0 \pmod{\mathfrak{p}_1^3}. \tag{27}$$

Since $b \equiv 0 \pmod 3$ we have $b^2 \equiv 1 \pmod 3$ or $b^2 \equiv 1 \pmod{\mathfrak{p}_1}$. Hence it follows from (27) that $m_1 \equiv m_1' \pmod{\mathfrak{p}_1}$, i.e., $m_1 \equiv m_1' \pmod 3$. We have thus shown that in a given field $m_1$ in (23) is fixed (mod 3) for any integer $\theta$ satisfying (22).

On the other hand, if $m_1, m_2, m_3, m_4$ are four arbitrary rational integers not divisible by 3 there exists a field $K$ containing an integer $\theta$ satisfying (22) such that the given $m$'s occur in (23). For let

$$f_i(x) = \phi_i^{(1)}(x) + 3(k_1\phi_i^{(1)}(x) + m_i)$$

where $m_i$ takes on the given values and $\phi_i^{(1)}(x)$, $i = 1, 2, 3, 4$, are the corresponding prime functions in (23). We may choose $k_i$ so that

$$k_i\phi_i^{(1)}(x) + m_i \not\equiv k_j\phi_j^{(1)}(x) + m_j$$

for $i \not\equiv j$. Then the discriminant of the polynomial $\Pi(x) = \Pi_{i=1}^4 f_i(x)$ is not

zero. Suppose it is divisible by exactly $3^\delta$. Then an equation of the form

$$f(x) = \Pi(x) + 3^{\delta+1}M(x) = 0$$

will determine the required field if $f(x)$ is irreducible. But we can obviously make $f(x)$ irreducible by choosing $M(x)$ so that $f(x)$ satisfies the Eisenstein irreducibility criterion with respect to a prime other than 3.

It follows from (24) that fields for which $m_1 \not\equiv m_2 \pmod 3$ have $\chi = 2$. On the other hand, if $m_1 \equiv m_2 \pmod 3$ and we choose $\theta$ so that $\mathfrak{p}_1^1 \cdot \mathfrak{p}_2^3 \mid \theta$, $\mathfrak{p}_3^1 \mid \theta+1$, and $\mathfrak{p}_4^1 \mid \theta+2$ it follows as in Theorem 5 that $\chi = 3$. We have thus proved the following theorem:

THEOREM 9. *There exist two types of fields of eighth degree in which*

$$[3] = \mathfrak{p}_1^2 \cdot \mathfrak{p}_2^2 \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4^2, \quad N\mathfrak{p}_i = 3.$$

*For fields of one type* $\chi = 2$, *for the other* $\chi = 3$.

It follows readily that

$$[3] = \mathfrak{p}_1^2 \cdot \mathfrak{p}_2^2 \cdot \mathfrak{p}_3^2 \cdot \mathfrak{p}_4^2 \cdot \mathfrak{p}_5, \quad N\mathfrak{p}_i = 3, i = 1,2,3,4, \quad N\mathfrak{p}_5 = 3^{n-8},$$

is an example of a type of ideal decomposition in the field of $n$th degree for which $\chi$ is not uniquely determined, i.e.,

THEOREM 10. *For fields of nth degree,* $n < 8$, $\chi$ *is not in general determined by the prime ideal decomposition.*

## IV. CONCLUSION

The complete solution of the problem of determining the possible values of $\chi$ for a given prime ideal decomposition is based on the solution of the problem of normalizing the factors (mod $p^a$) of the characteristic equation. The latter problem is a fundamental problem in the theory of algebraic numbers. In the theory of Kronecker a common index divisor must divide the index form, a homogeneous form in $n$ variables of degree $\frac{1}{2}n(n-1)$, for all integral rational values of the variables. By using known theorems on the greatest common divisor of integers represented by forms it can be shown that the greatest common index divisor of a field of $n$th degree is not greater than $(\frac{1}{2}n(n-1))!$. The results of this paper seem to indicate that the expression for $\chi$ in Theorem 3 is the best maximum for $\chi$ for a field of $n$th degree, but a proof seems to revert to the more fundamental problem mentioned above.

YALE UNIVERSITY,
NEW HAVEN, CONN.

# THE BOUNDARY PROBLEM ASSOCIATED WITH A DIFFERENTIAL SYSTEM RATIONAL IN THE PARAMETER*

BY

RUDOLPH E. LANGER

1. **Introduction.** It is the purpose of this paper to consider the boundary problem associated either with the ordinary differential equation

$$u''(x) + P_1(x,\lambda)u'(x) + P_2(x,\lambda)u(x) = 0, \tag{1}$$

or with the vector differential equation of the second order

$$U'(x)\cdot - \mathfrak{Q}(x,\lambda)\, U(x)\cdot = \mathcal{O}, \tag{2}$$

when the coefficients involved both in the equation and the adjoined boundary conditions are suitably restricted rational functions of the parameter. For the more familiar case in which the parameter enters the system in polynomial form an extensive theory is at hand, and this theory, together with the available asymptotic forms of solutions of the equation, is of direct application in the present discussion.

The extension to the case of coefficients rational in $\lambda$ for the case of an equation of the first order has previously been made†, and that simpler discussion is typical of the present one in that the differential system is found, very naturally, to exhibit in the neighborhood of a pole of the coefficients the properties which are peculiar to the vicinity of $\lambda = \infty$ when the parameter occurs in the form of polynomials. Thus, for instance, any neighborhood of such a pole is found with proper hypotheses to contain infinitely many characteristic values. Roughly speaking, these values are, therefore, more numerous than in the classical case, and accountable to this fact a single determination of coefficients is found to be possible to serve formally for the simultaneous expansion of several "arbitrary" functions in series of solutions of the system.

The present paper is based extensively upon a recent paper by the author‡ devoted to the case in which the coefficients are $\lambda$-polynomials. In fact the deductions are as a rule not carried beyond a point at which the facts estab-

* Presented to the Society, September 7, 1928; received by the editors August 2, 1929.

† Langer, R. E., these Transactions, vol. 25 (1923), p. 155.

‡ Langer, R. E., these Transactions, vol. 31, pp. 868–906. This paper will be referred to in the text by the designation L. The reader is referred to this paper for references to the literature.

lished in the paper cited become available. The vector form of the differential system is first shown to be the more general and is then made the primary subject of consideration.

2. **The differential equation.** The basic hypothesis on the structure of the equation (1) with which we are concerned is that the coefficients $P_j(x, \lambda)$, $j=1, 2$, be rational functions of $\lambda$ with poles independent of $x$ and of order $j$. If these poles are designated in the complex $\lambda$ plane by $a_h$, $h=1, 2, \cdots, m$,* the explicit form of the coefficients is

$$
\begin{aligned}
P_1(x,\lambda) &= p_{10}(x) + \sum_{h=1}^{m} \frac{\overset{h}{p}_{11}(x)}{\lambda - a_h}, \\
P_2(x,\lambda) &= p_{20}(x) + \sum_{h=1}^{m} \left\{ \frac{\overset{h}{p}_{22}(x)}{(\lambda - a_h)^2} + \frac{\overset{h}{p}_{21}(x)}{\lambda - a_h} \right\}.
\end{aligned} \tag{3}
$$

A differential equation of form (1), (3) may always be obtained from the system of two equations of first order

$$
\begin{aligned}
u_1'(x) &= q_{11}(x,\lambda)u_1(x) + u_2(x), \\
u_2'(x) &= q_{21}(x,\lambda)u_1(x) + q_{22}(x,\lambda)u_2(x),
\end{aligned} \tag{4}
$$

if the coefficients $q_{ij}(x, \lambda)$ are rational in $\lambda$ with poles of the first order at the points $a_h$. In fact if we set

$$
q_{ij}(x,\lambda) = \overset{0}{q}_{ij}(x) + \sum_{h=1}^{m} \frac{\overset{h}{q}_{ij}(x)}{\lambda - a_h}, \tag{5}
$$

the elimination of $u_2(x)$ from the system (4) yields for $u_1(x)$ an equation of form (1), (3) with coefficients which may be computed from the formulas

$$
\begin{aligned}
&\text{(a)} && \overset{h}{q}_{11} + \overset{h}{q}_{22} = -\overset{h}{p}_{11} && (h = 1,2, \cdots, m), \\
&\text{(b)} && \overset{0}{q}_{11} + \overset{0}{q}_{22} = -p_{10}, && \\
&\text{(c)} && \overset{h}{q}_{11}\overset{h}{q}_{22} = \overset{h}{p}_{22} && (h = 1,2, \cdots, m), \\
&\text{(d)} && \overset{h}{q}_{22}\overset{0}{q}_{11} + \overset{h}{q}_{11}\overset{0}{q}_{22} - \overset{h}{q}_{21} + \sum_{l \neq h} \frac{\overset{h}{q}_{11}\overset{l}{q}_{22} - \overset{l}{q}_{11}\overset{h}{q}_{22}}{a_h - a_l} - (\overset{h}{q}_{11})' = \overset{h}{p}_{21} && \\
& && && (h = 1,2, \cdots, m), \\
&\text{(e)} && \overset{0}{q}_{11}\overset{0}{q}_{22} - (\overset{0}{q}_{11})' - \overset{0}{q}_{21} = p_{20}. &&
\end{aligned} \tag{6}
$$

We wish to show conversely that the equation (1), (3) is always expres-

* For the sake of symmetry we assume the poles to be all located in the finite portion of the plane. This involves no loss of generality since it can always be achieved by at most a change of parameter.

sible in the form of a system (4), (5). This is readily done by establishing as follows the solvability of the system of equations (6) for the quantities $q$ with given values $p$. To begin with the equations (6a) and (6c), determine the quantities $q_{11}^h$, $q_{22}^h$ for each fixed value of $h$ as the roots of the equation

$$(q^h)^2 + p_{11}^h(q^h) + p_{22}^h = 0.$$

In equations (6b) and (6d) the unknowns are, then, the $(m+2)$ quantities $q_{11}^0$, $q_{22}^0$ and $q_{21}^h$, $h=1, 2, \cdots, m$, and in these quantities the equations are linear and only $(m+1)$ in number. Lastly the quantity $q_{21}^0$ is determined by the equation (6e). It is clear, therefore, since the system (4) is a restricted case of the vector equation, that the theory of the equation (1) is included in that of the equation (2), the elements of the coefficient matrix $\mathfrak{Q}(x, \lambda)$ being given in structure by the formula (5). We shall confine our considerations, therefore, to this latter equation, writing it in the more explicit form

$$U'(x)\cdot - \left\{ \sum_{h=1}^{m} \frac{1}{\lambda - a_h} \mathfrak{Q}^h(x) + \mathfrak{Q}^0(x) \right\} U(x)\cdot = \mathcal{O}. \tag{7}$$

By way of hypotheses the elements of the coefficient matrices in (7) will be assumed to possess such derivatives as the following deductions require, and as an essential feature we shall suppose further that for each value of $h$ the roots $\theta_1^h(x)$ $\theta_2^h(x)$ of the equation

$$\begin{vmatrix} q_{11}^h(x) - \theta^h & q_{12}^h(x) \\ q_{21}^h(x) & q_{22}^h(x) - \theta^h \end{vmatrix} = 0$$

are distinct, non-vanishing and of constant arguments. If these arguments are suitably determined we have then

$$\begin{gathered} \theta_1^h(x) \not\equiv \theta_2^h(x), \\ \theta_j^h(x) = |\theta_j^h(x)|\, e^{ic_j^h};\ j = 1,2;\ |c_1^h - c_2^h| \leqq \pi. \end{gathered} \tag{8}$$

Without loss of generality we may consider the variable $x$ to range over the interval (0, 1).

3. **The related polynomial equations.** The change of parameter

$$\rho_k = \frac{\beta_k}{\lambda - a_k}, \tag{9}$$

in which $\beta_k$ is any constant, projects the pole $\lambda = a_k$ to $\rho_k = \infty$ and gives to the equation (7) the form

$$U'\cdot - \left\{ \beta_k^{-1} \mathfrak{Q}^k \rho_k + \left( \mathfrak{Q}^0 + \sum_{h \neq k} a_{kh}^{-1} \mathfrak{Q}^h \right) - \sum_{h \neq k} \frac{\beta_k}{a_{kh}(a_{kh}\rho + \beta_k)} \mathfrak{Q}^h \right\} U\cdot = \mathcal{O},$$

where we have abbreviated by setting

$$a_k - a_h = a_{kh}.$$

This equation we shall further transform by the change of dependent variable

$$U(x)\cdot = \Phi_k(x)\Upsilon^k(x)\cdot, \tag{10}$$

choosing the matrix $\Phi_k(x)$, as is always possible, L§3, so that in the resulting form of the equation, i.e.,

$$\Upsilon'^k(x)\cdot - \left\{\mathcal{R}^k(x)\rho_k + \mathcal{B}^k(x) + \sum_{h\neq k}\frac{1}{a_{kh}\rho_k + \beta_k}\mathcal{B}^{kh}(x)\right\}\Upsilon^k(x)\cdot = \mathcal{O}, \tag{11}$$

the leading coefficient matrices have the structure

$$\mathcal{R}^k(x) = \begin{pmatrix} r_1{}^k(x) & 0 \\ 0 & r_2{}^k(x) \end{pmatrix}, \quad \mathcal{B}^k(x) = \begin{pmatrix} 0 & b_1{}^k(x) \\ b_2{}^k(x) & 0 \end{pmatrix},$$

with

$$r_j(x) = \theta_j(x)/\beta_k. \tag{12}$$

From the hypotheses (8) and relation (12), it follows that the quantities $r_j(x)$ are distinct from each other and from zero, and that they are of constant arguments depending upon the value $\beta_k$. The particular choice

$$\beta_k = \exp\left(\frac{i}{2}(c_1{}^k + c_2{}^k)\right)$$

which we shall now make is found directly then to lead to the following relations valid for $x$ on the interval (0, 1), i.e.,

$$\begin{aligned} &r_1{}^k(x) \not\equiv r_2{}^k(x), \\ &r_1{}^k(x) = |\, r_1{}^k(x)\,|\, e^{i\alpha_k} \\ &r_2{}^k(x) = |\, r_2{}^k(x)\,|\, e^{-i\alpha_k}, \quad 0 \leqq \alpha_k \leqq \pi/2. \end{aligned} \tag{13}$$

It will frequently be found convenient to use the given equation in the form (11) as well as in the form (7), and for this reason we note here in particular the relation

$$\mathcal{R}^k(x) = \frac{1}{\beta_k}\Phi_k{}^{-1}(x)\mathcal{Q}^k(x)\Phi_k(x). \tag{14}$$

The polynomial equation

$$\bar{\Upsilon}'^k(x)\cdot - \{\mathcal{R}^k(x)\rho_k + \mathcal{B}^k(x)\}\bar{\Upsilon}^k(x)\cdot = \mathcal{O}, \tag{15}$$

derived from equation (11) by retaining only the principal part of the coefficient matrix at $\rho_k = \infty$, is in so-called normal form, L§3, due to the relations (13). We shall call this equation the $k$th related polynomial equation of the equation given. It is clear that the procedure outlined above for the pole $a_k$ may be made to yield for the given equation a unique related normal polynomial equation associated with each of the poles $a_h$, $h=1, 2, \cdots, m$.

The equation adjoint to (7) may be written

$$\cdot U'(x) + \cdot U(x)\left\{\sum_{h=1}^{m}\frac{1}{\lambda - a_h}\mathcal{Q}^h(x) + \mathcal{Q}^0(x)\right\} = \mathcal{O}, \tag{16}$$

and this is reduced by the transformation

$$\cdot U(x) = \cdot Z^k(x)\Phi_k^{-1}(x),$$

and the substitution (9) to the alternative form

$$\cdot Z'^k(x) + \cdot Z^k(x)\left\{\mathcal{R}^k(x)\rho_k + \mathcal{B}^k(x) + \sum_{h\neq k}\frac{1}{a_{kh}\rho_k + \beta_k}\mathcal{B}^{kh}(x)\right\} = \mathcal{O}. \tag{17}$$

For this equation the $k$th related polynomial equation is evidently

$$\cdot \bar{Z}'^k(x) + \cdot \bar{Z}^k(x)\{\mathcal{R}^k(x)\rho_k + \mathcal{B}^k(x)\} = \mathcal{O}.$$

4. **The asymptotic solutions.** The asymptotic forms of a pair of solutions of the polynomial equation (15), valid when $\rho_k$ is confined to any region in which the quantity

$$\text{Re}\,(\rho_k\{r_1^k(x) - r_2^k(x)\})^*$$

remains either greater than or less than any specific constant and $|\rho_k|$ is sufficiently large, are given, L§4, by the columns of a matrix

$$\bar{Y}^k(x) = \bar{\mathcal{P}}^k(x,\rho_k)\,\mathcal{E}^k(x,\rho_k), \tag{18a}$$

with

$$\begin{aligned} e_{ij}^k(x,\rho) &= \delta_{ij}e^{\rho_k R_j^k(x)}, \\ \bar{p}_{ij}^k(x,\rho) &= \rho_k^{\delta_{ij}-1}[\bar{p}_{ij}^k(x)], \quad \bar{p}_{jj}^k(x) \equiv 1, \end{aligned} \tag{18b}$$

the notation being that of the paper cited. A review of the reasoning and manipulation by which these forms are deduced† shows directly that the process requires no essential modification if the considerations are transferred

* Re $(\omega)$ is used to denote "the real part of $\omega$."

† Birkhoff and Langer, Proceedings of the American Academy of Arts and Sciences, vol. 58 (1923), p. 51, Sections V and VI.

from the polynomial equation to the equation (11), the terms in $(a_{kh}\rho_k+\beta_k)^{-1}$ being thought of as first expanded in powers of $1/\rho$.* Moreover, the determination of the initial terms of the asymptotic series is found to be independent of the matrices $\mathcal{B}^{kh}(x)$, and in consequence the equation (15) may be shown to admit of a pair of solutions given asymptotically by the columns of a matrix

$$\mathcal{Y}^k(x) = \mathcal{P}^k(x,\rho_k)\mathcal{E}^k(x,\rho_k), \tag{19}$$

in which the formulas (18b) may be used for $\mathcal{E}^k$ and $\mathcal{P}^k$ respectively. In entirely similar fashion the rows of a matrix

$$\mathcal{Z}^k(x) = (\mathcal{E}^k(x,\rho_k))^{-1}\mathcal{M}(x,\rho_k),$$

with

$$m^k_{ij}(x,\rho_k) = \rho_k^{s_{ij}-1}[m^k_{ij}(x)], \quad m^k_{jj}(x) \equiv 1,$$

are found to give the asymptotic forms of a pair of solutions of the adjoint equation (17) or the polynomial equation associated with it.

To insure the availability of these forms over the entire remote portion of the $\rho_k$ plane we must make the hypothesis that

$$\arg\{r_1^k(x) - r_2^k(x)\} = A^k \quad \text{(a constant)},$$

for each value of $k$. The asymptotic forms are applicable then for $\rho_k$ restricted to any half-plane bounded by a line parallel to the ray

$$\arg \rho_k = \pi/2 - A^k.$$

5. **The boundary condition and the characteristic values.** As the boundary condition to be imposed upon the given equation (7) we prescribe the relation

$$\mathcal{W}_0(\lambda)\, U(0)\cdot + \mathcal{W}_1(\lambda)\, U(1)\cdot = \mathcal{O}, \tag{20}$$

the elements of the coefficient matrices $\mathcal{W}_0$ and $\mathcal{W}_1$ being any rational functions of $\lambda$ with poles only in the points $a_h$. If we suppose the elements of these matrices to be reduced to a common denominator, this denominator is of the form

$$\prod_{h=1}^{m}(\lambda - a_h)^{l_h},$$

in which each exponent $l_h$ is a positive integer or zero. Since by equation (7)

* If equation (14) is one which is reducible to a single equation of the second order, the coefficients being written as series in powers of $1/\rho$, the asymptotic forms are given directly by Birkhoff, these Transactions, vol. 9 (1908), and by Tamarkin, Mathematische Zeitschrift, vol. 27 (1927), p. 1 (and Petrograd, 1917).

$\lambda$ is restricted from assuming the values $a_h$ we may, for any value of $k$, $k=1, 2, \cdots, m$, multiply the equation (20) by the quantity

$$\prod_{l \neq k} \left(\frac{\lambda - a_h}{\lambda - a_k}\right)^{l_h},$$

without thereby altering it in content. The elements of the new coefficient matrices, however, have poles only at the point $a_k$, and under the change of parameter (9) and the substitution (10) the condition takes the form

$$\mathcal{W}^k_{12}(\rho_k)\Upsilon^k(0)\cdot + \mathcal{W}^k_{34}(\rho_k)\Upsilon^k(1)\cdot = \mathcal{O}. \tag{21}$$

In this the elements of the matrices $\mathcal{W}^k_{jl}(\rho_k)$ are polynomials in $\rho_k$. The values of $\lambda$, if any, which satisfy the system (7), (20) clearly correspond under the relation (9) to the values of $\rho_k$ which satisfy the system (11), (21) and vice versa.

The equation (11) was found in §4 to admit of solutions which for $|\rho_k|$ sufficiently large and $\rho_k$ confined to a suitable half-plane are represented asymptotically by the columns of the matrix (19). By familiar reasoning then, L §6, the characteristic values of the given boundary problem are shown to be the roots of the determinant equation

$$|\mathcal{W}^k_{12}(\rho_k)\mathcal{P}^k(0,\rho_k) + \mathcal{W}^k_{34}(\rho_k)\mathcal{P}^k(1,\rho_k)\,\mathcal{E}^k(1,\rho_k)| = 0. \tag{22}$$

We prescribe now for the associated polynomial equation (15) the boundary condition

$$\overline{\mathcal{W}}^k_{12}(\rho_k)\overline{\Upsilon}^k(0)\cdot + \overline{\mathcal{W}}^k_{34}(\rho_k)\overline{\Upsilon}^k(1)\cdot = \mathcal{O}, \tag{23}$$

in which the coefficient matrices are to be determined as finite sums of powers of $\rho_k$ to satisfy relations of the type

$$\begin{aligned}
\overline{\mathcal{W}}^k_{12}(\rho_k) &= \mathcal{W}^k_{12}(\rho_k)\mathcal{P}^k(0,\rho_k)(\overline{\mathcal{P}}^k(0,\rho_k))^{-1} + \frac{1}{\rho^s}\Omega^k_{12}(\rho_k), \\
\overline{\mathcal{W}}^k_{34}(\rho_k) &= \mathcal{W}^k_{34}(\rho_k)\mathcal{P}^k(1,\rho_k)(\overline{\mathcal{P}}^k(1,\rho_k))^{-1} + \frac{1}{\rho^s}\Omega^k_{34}(\rho_k).
\end{aligned} \tag{24}$$

In these relations the matrices $\Omega_{jl}$ are to be series of negative powers of $\rho_k$ and $s$ is to be an integer which will be specified later. By at most the multiplication of (23) by a power of $\rho_k$ the coefficient matrices can be made polynomials. We shall designate the system (15), (23) as the $k$th associated polynomial system for the given system (7), (20).

The characteristic values of the boundary problem (15), (23) are the roots of the equation

$$|\overline{\mathcal{W}}_{12}^{k}(\rho_k)\overline{\mathcal{P}}^{k}(0,\rho_k) + \overline{\mathcal{W}}_{34}^{k}(\rho_k)\overline{\mathcal{P}}^{k}(1,\rho_k)\ \mathcal{E}^{k}(1,\rho_k)| = 0,$$

and on expansion of the determinant this equation takes the more explicit form, L §6,

$$D_0^k(\rho_k) + D_1^k(\rho_k)e^{\rho_k R_1^k} + D_2^k(\rho_k)e^{\rho_k R_2^k} + D_3^k(\rho_k)e^{\rho_k(R_1^k+R_2^k)} = 0, \tag{25}$$

the coefficients being of the structure

$$D_r^k(\rho_k) = \rho_k^{\kappa_r^k}[\alpha_r^k], \quad \alpha_r^k \neq 0. \tag{26}$$

Now because of (24) the expansion of the equation (22) will be precisely of the form (25), (26) provided the integer $s$ in (24) is chosen sufficiently large. We suppose such a choice of $s$ to have been made. Inasmuch as the characteristic equations of the given and associated systems are then both of the form (25)* it follows that the characteristic values of the given boundary problem which lie in the remote portion of the $\rho_k$ plane are represented asymptotically by those of the associated polynomial problem in $\rho_k$. From this we conclude in particular, L §7, that if there exist in the remote $\rho_k$ plane enumerably many characteristic values, their distribution is such that a sequence of successively larger contours $C_n^k$, $n=n_1, n_2, n_3, \cdots$, may be drawn in the $\rho_k$ plane to pass between the characteristic values and to have the following properties: (a) on the contours of the set the quotient of the left member of equation (25) by any one of its terms is uniformly bounded from zero; (b) in the annular region bounded by a pair of the contours, $C_{n_h}^h$ and $C_{n_l}^l$ there lie just $|n_h - n_l|$ characteristic values.

We shall denote by $\rho_{ks}$ the characteristic values for which $|\rho_k|$ is large, assigning the subscripts so that

$$|\rho_{ks}| \leqq |\rho_{k,s+1}|.$$

By means of the substitution (9) the corresponding characteristic values of the system (7), (20) which lie in a neighborhood of $\lambda = a_k$ are obtained. These values have the point $a_k$ as a limit point. We denote them by $\lambda_{ks}$. The substitution (9) similarly transforms the contours $C_n^k$ into a sequence of decreasing contours $C_n^k$ surrounding and converging to the point $a_k$. We shall find these contours of especial usefulness in the subsequent developments.

If we consider that in the discussion just concluded $k$ may have any of the values $1, 2, \cdots, m$, it follows that if each of the associated polynomial

* I.e., they differ at most in the later terms of the quantities $[\alpha_r^k]$.

systems admits of infinitely many characteristic values, then the given boundary problem also admits of infinitely many characteristic values, these values having limit points at and only at the poles of the coefficients. In proceeding we shall suppose that the boundary problem in hand is one in which this situation obtains.

The characteristic solutions of the given equation in the forms (7) and (11) for the value $\lambda=\lambda_{ks}$ we shall denote by $U^{ks}(x)\cdot$ and $Y^{ks}(x)\cdot$ respectively. In like manner the symbols $\cdot\,\mathcal{U}^{ks}(x)$ and $\cdot\,\mathcal{Z}^{ks}(x)$ shall designate the characteristic solutions of the adjoint equations.

6. **The formal expansion of arbitrary vectors.** The characteristic values, as shown in the preceding section, have as limit points the $m$ poles of the coefficients. They are, therefore, roughly speaking $m$ times as numerous as in the classical polynomial case, and because of this we shall seek to expand formally not merely a single vector but a set of $m$ vectors arbitrarily given. Thus if the given vectors are denoted by $\mathfrak{F}^h(x)\cdot$, $h=1, 2, \cdots, m$, we shall seek a set of constant coefficients $c_{rs}$ to serve simultaneously in the $m$ series developments

$$\mathfrak{F}^h(x)\cdot = \sum_{r=1,s=1,2,\cdots,m}^{\infty} \frac{c_{rs}}{\lambda_{rs}-a_h}\, U^{rs}(x)\cdot \qquad (h=1,2,\cdots,m). \tag{27}$$

The determination of these coefficients requires but little modification of the procedure applied in L §8. Thus the equations (7) and (16) lead by a familiar procedure to the relation

$$\int_0^1 \cdot\,\mathcal{U}^{pq} \sum_{h=1}^{m} \frac{1}{(\lambda_{rs}-a_h)(\lambda_{pq}-a_h)} \mathcal{Q}^h U^{rs}\cdot dx + \frac{1}{\lambda_{rs}-\lambda_{pq}}\{\cdot\,\mathcal{U}^{pq}U^{rs}\cdot\,\Big|_0^1\} = \mathcal{O}, \quad \text{for} \quad (p,q)\neq(r,s). \tag{28}$$

If we suppose then that the $c_{rs}$ are in addition to the relations (27) also amenable to a relation of the form

$$\sum_{r=1,s=1,2,\cdots,m}^{\infty} \frac{c_{rs}}{\lambda_{rs}-\lambda_{pq}}\cdot\,\mathcal{U}^{pq}U^{rs}\cdot\,\Big|_0^1 = \{\cdot\,\mathcal{U}^{pq}(1)\mathcal{C}_1 + \cdot\,\mathcal{U}^{pq}(0)\mathcal{C}_0\}K(\lambda_{pq})\cdot,$$

in which the $\mathcal{C}_j$ are constant matrices and $K(\lambda_{pq})\cdot$ is a vector whose specification we defer, we find directly on multiplying the left member of (28) by $c_{rs}$ and summing over all values of $r$ and $s$, the equality

$$(29)\quad \int_0^1 \cdot\, U^{pq} \sum_{h=1}^{m} \frac{1}{\lambda_{pq} - a_h} \mathcal{Q}^h \mathfrak{F}^h \cdot dx + \{ \cdot\, U^{pq}(1)\mathcal{C}_1 + \cdot\, U^{pq}(0)\mathcal{C}_0 \} K(\lambda_{pq}) \cdot$$
$$= c_{pq} \Big\{ \int_0^1 \cdot\, U^{pq} \sum_{h=1}^{m} \frac{1}{(\lambda_{pq} - a_h)^2} \mathcal{Q}^h\, U^{pq} \cdot dx + [(\lambda_{rs} - \lambda_{pq})^{-1} \cdot\, U^{pq}\, U^{rs} \cdot \big|_0^1]_{(r,s)=(p,q)} \Big\} \cdot$$

Inasmuch as each member of this equation is a matrix in which all elements are the same the relation may be looked upon as formally determining the coefficient $c_{pq}$.

As in L §9 we may utilize the Green's function $\mathcal{G}(x, t, \lambda)$ of the given system (7), (20) to derive from the relation (29) a set of contour integrals associated with the formal series (27). Thus we are led by reasoning analogous to that of the paper cited to the relation

$$\frac{c_{pq}}{\lambda_{pq} - a_k} U^{pq}(x) \cdot = \int_0^1 \mathcal{G}^{pq}(x,t) \sum_{h=1}^{m} \frac{1}{(\lambda_{pq} - a_h)(\lambda_{pq} - a_k)} \mathcal{Q}^h(t) \mathfrak{F}^h(t) \cdot dt$$
$$+ \frac{1}{\lambda_{pq} - a_k} \{ \mathcal{G}^{pq}(x,1)\mathcal{C}_1 + \mathcal{G}^{pq}(x,0)\mathcal{C}_0 \} K(\lambda_{pq}) \cdot \cdot$$

Further manipulations lead to the association of quantities

$$(30)\qquad \sum \frac{c_{pq}}{\lambda_{pq} - a_k} U^{pq}(x) \cdot \sim \mathfrak{I}_N(x, \cdot \mathfrak{F}^k \cdot),$$

where

$$(31)\quad \mathfrak{I}_N(x, \mathfrak{F}^k \cdot) = \mathcal{A}^k(x) \cdot$$
$$+ \frac{1}{2\pi i} \int_0^1 \int_{C_N} \mathcal{G}(x,t,\lambda) \sum_{h=1}^{m} \frac{1}{(\lambda - a_h)(\lambda - a_k)} \mathcal{Q}^h(t) \bar{\mathfrak{F}}^h(t) \cdot d\lambda dt.$$

In these relations the sum in (30) is to extend over those and only those values of $p$ and $q$ for which the point $\lambda_{pq}$ is enclosed by the contour $C_N$, and in (31)

$$\mathcal{A}^k(x) \cdot = x\mathfrak{F}^k(1) \cdot + (1 - x)\mathfrak{F}^k(0) \cdot,$$
$$\bar{\mathfrak{F}}^k(x) \cdot = \mathfrak{F}^k(x) \cdot - \mathcal{A}^k(x) \cdot .$$

The quantity (31) is the analogue of the quantity L (60). Its convergence as $C_N$ is replaced successively by contours enclosing a greater and greater number of characteristic values corresponds to the convergence of the $k$th development (27).

A particular choice of contours $C_N$ which is especially useful is that given by the sequence

$$C_N = -c_n^1 - c_n^2 - \cdots - c_n^m,$$

where the contours $c_n^k$ on the right are those of the sequences described in §5.* By this definition $C_N$ encloses all those characteristic values which are not enclosed by any of the $c_n^k$. As $n \to \infty$ these latter contours converge toward the respective poles $a_k$, and the corresponding contours $C_N$ accordingly enclose more and more of the characteristic values. We shall suppose in the following discussion that this choice of the contours $C_N$ has been made.

A set of quantities $\mathcal{J}_N(x, \mathfrak{F}^k \cdot)$, analogous to those given in L (65), whose convergence corresponds formally to the summability of the developments (27) may also be deduced as in L §10. Thus if we consider the contour $C_n^k$ of §5 to be a circle with radius $\Lambda_{kn}$, we may define the matrices

$$\mathfrak{S}_{kn}(\lambda) = \left(\delta_{ij}\left\{1 - \left[\frac{\beta_k}{\Lambda_{kn}(\lambda - a_k)}\right]^{\mu_k}\right\}^{\sigma_{jk}}\right), \tag{32}$$

in which the $\sigma_{jk}$ are constants positive or zero and $\mu_k$ is determined as in the $k$th associated polynomial case L §13. The quantities in question may then be written as

$$\mathcal{J}_N(x, \mathfrak{F}^k \cdot) = \mathcal{A}^k(x) \cdot + \sum_{l=1}^{m} \frac{1}{2\pi i} \int_0^1 \int_{-c_n^l} \Phi_l(x) \mathfrak{S}_{ln}(\lambda) \Phi_l^{-1}(x)\, \mathcal{G}(x, t, \lambda) \cdot \sum_{k=1}^{m} \frac{1}{(\lambda - a_k)(\lambda - a_k)} \mathcal{Q}^k(t) \bar{\mathfrak{F}}^k(t) \cdot d\lambda dt. \tag{33}$$

It is clear that for $\sigma_{jk} = 0$, $j = 1, 2$; $k = 1, 2, \cdots, m$, the matrices (32) reduce to the unit matrix and the quantities (33) reduce to those given by formula (31).

7. **The convergence of $\mathcal{J}_N(x, \mathfrak{F}^k \cdot)$.** Let the left members of the equations (7) and (11) be designated respectively by $\mathcal{L}(U \cdot, \lambda)$ and $\mathcal{L}_k(\Upsilon^k \cdot \rho_k)$. Then the equation

$$\mathcal{L}(U \cdot, \lambda) = \Psi(x) \cdot, \tag{34}$$

with any choice of the vector on the right, is found to be equivalent under the substitutions (9) and (10) to the equation

$$\mathcal{L}_k(\Upsilon^k \cdot, \rho_k) = \Phi_k^{-1}(x) \Psi(x) \cdot. \tag{34a}$$

* It is only as a matter of convenience that we write here the same subscript for each of the contours $c_n^k$.

Now in terms of the Green's function as defined above the solution of equation (34) with condition (20) is given by the formula

$$U(x)\cdot = \int_0^1 \mathcal{G}(x,t,\lambda)\Psi(t)\cdot dt, \tag{35}$$

and if we similarly define $\mathcal{G}_k(x, t, \lambda)$ as the Green's function of the system (11), (21), the solution of (34a) with the condition (21) is

$$\Upsilon^k(x)\cdot = \int_0^1 \mathcal{G}^k(x,t,\rho_k)\Phi_k^{-1}(t)\Psi(t)\cdot dt. \tag{35a}$$

From (35) and (35a) we conclude, therefore, that

$$\mathcal{G}(x,t,\lambda) \equiv \Phi_k(x)\,\mathcal{G}_k(x,t,\rho_k)\Phi_k^{-1}(t) \qquad (k = 1,2,\cdots,m).$$

With this formula we shall transform the quantity (33) by applying on each contour $c_n^l$ the respective transformation (9). In this manner we obtain in virtue of (14)

$$\begin{aligned}\mathcal{J}_N(x,\mathfrak{F}^k\cdot) &= \mathcal{A}^k(x)\cdot \\ &- \frac{\Phi_k(x)}{2\pi i}\int_0^1\int_{c_n^k} \mathfrak{S}_{kn}\left(a_k+\frac{\beta_k}{\rho_k}\right)\mathcal{G}_k(x,t,\rho_k)\mathcal{R}^k(t)\Phi_k^{-1}(t)\bar{\mathfrak{F}}^k(t)\cdot d\rho_k dt \\ &- \sum_{(l,h)\neq(k,k)} \frac{\Phi_l(x)}{2\pi i}\int_0^1\int_{c_n^l} \mathfrak{S}_{ln}\left(a_l+\frac{\beta_l}{\rho_l}\right)\mathcal{G}_l(x,t,\rho_l)\Phi_l^{-1}(t)\mathcal{Q}^h(t)\bar{\mathfrak{F}}^h(t)\cdot \\ &\qquad \cdot\frac{\beta_l d\rho_l}{(a_{lh}\rho_l+\beta_l)(a_{lk}\rho_l+\beta_l)}dt.\end{aligned} \tag{36}$$

Consider in this formula the Green's functions occurring in the integrands. From the explicit formula, L (73), it is evident that when $|\rho_h|$ is large $\mathcal{G}_h(x, t, \rho_h)$ is asymptotically given by the Green's function of the $h$th associated polynomial system. Since the matrix $\mathfrak{S}_{hn}(a_h+\beta_h/\rho_h)$ is the same in form as the matrix L (63), it follows that the first of the integrals in (36) is precisely of the structure of the integral in L (65). The integrand is, therefore, asymptotically the same as $\Phi_k(x)$ times the integrand obtained from the expansion of $\Phi_k^{-1}(x)\mathfrak{F}^k(x)\cdot$ with regard to the $k$th associated polynomial system, and the convergence of the integral in the latter case implies the convergence of the integral in hand under the same hypotheses.

The integrals occurring under the sign of summation in (36) resemble that just considered with the exception that a factor

$$\frac{1}{(a_{lh}\rho_l+\beta_l)(a_{lk}\rho_l+\beta_l)}$$

occurs in the integrand. Inasmuch as $h$, $k$, and $l$ are not all equal, this factor is of at least the degree one in $1/\rho$, and it is clear that the convergence of the integral already considered, for each choice of $k$, to any finite value implies the convergence of the integrals in hand to the value $\mathcal{O}$. We are led, therefore, to state the following

THEOREM. *The set of expressions* $\mathcal{J}_N(x, \mathfrak{F}^k \cdot)$, $k = 1, 2, \cdots, m$, *is equiconvergent with the corresponding set of expressions,* L (65), *obtained from the respective expansion of the kth vector of the set* $\mathfrak{F}^h(x) \cdot$, $h = 1, 2, \cdots, m$, *with regard to the kth associated polynomial system.*

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

# INVOLUTORIAL TRANSFORMATIONS IN $S_3$ OF ORDER $n$ WITH AN $(n-1)$-FOLD LINE*

BY

LEAMAN A. DYE

1. **Introduction.** Montesano† has given a brief synthetic discussion of the existence of involutorial transformations $I_n$ of order $n$ with an $(n-1)$-fold line $l$. He showed that the planes through $l$ are interchanged in pairs by $I_n$ and that the lines in one plane are transformed into lines in the conjugate plane. He also showed that the $I_n$ could be defined by the aid of two curves of order $n-1$ situated on the fundamental surface $F_{n-1}:l^{n-2}$ which is the image of $l$.

In this paper the $F_{n-1}:l^{n-2}$ and an $F_n:l^{n-1}$ are used to define an involutorial transformation of order $2n-1$ with a $(2n-3)$-fold line which, if certain conditions are satisfied, reduces to an involutorial transformation of order $n$ with an $(n-1)$-fold line. The explicit analytical forms of $I_3$ and $I_4$ are found by this method. For larger values of $n$ it is convenient to define $I_n$ by other means. There is a net of surfaces of order $m$, $m+1$, or $m+2$ according as $n=3m-1$, $3m$, or $3m+1$, which is transformed by $I_n$ into a net of surfaces of order $m$. These nets are used to define the involutorial transformation and the equations of $I_5$, $I_6$, and $I_7$ are derived. A method is given for mapping $I_n$ on ordinary space so that it is apparent that $I_n$ is rational.

2. **The birational transformation of type $(n, n)$ with an $(n-1)$-fold line.** Two surfaces $F_n$ of order $n$ having an $(n-1)$-fold line $l$ in common, meet in a residual curve $C_{2n-1}$. Any plane through $l$ meets each $F_n$ in a residual line, and the two lines meet in a point on the $C_{2n-1}$. Hence the $C_{2n-1}$ meets $l$ in $2n-2$ points. If the two surfaces have a $C_n$ in common meeting $l$ in $n-1$ points, then by Noether's‡ formulas the $C_{2n-1}$ will consist of the $C_n$ and $n-1$ lines $l_i$ meeting $l$.§ Conversely through $n-1$ lines $l_i$ meeting $l$, pass $\infty^{n+2}$ surfaces $F^n:l^{n-1}$, $\Sigma l_i$ such that any two meet in a $C_n$ which meets the line $l$ in $n-1$ points and each line $l_i$ in one point. Three of these surfaces meet in $n$ points and if we fix $n-1$ of the points, we have a homoloidal web of surfaces $F_n:l^{n-1}$, $\Sigma l_i$, $\Sigma P_i$.

* Presented to the Society, February 22, 1930; received by the editors January 20, 1930.

† D. Montesano, *Su una classe di trasformazioni involutorie dello spazio*, Istituto Lombardo, Rendiconti, (2), vol. 21 (1888), pp. 688–690.

‡ M. Noether, *Sulle curve multiple di superficie algebriche*, Annali di Matematica, (2), vol. 5 (1871), pp. 163–177.

§ H. P. Hudson, *Cremona Transformations in Plane and Space*, p. 316.

There exists therefore a Cremona transformation $T_{n,n}$ between two spaces $(x)$ and $(x')$ by which a plane in $(x') \sim$ a surface of the web of $F_n$ in $(x)$. A plane and a surface $F_n$ in the $(x)$ space $\sim$ a surface $F_n'$ and a plane in the $(x')$ space. The curves of intersection correspond and have the same genus.* Hence a plane section of $F_n'$ is a $C_n'$ with an $(n-1)$-fold point and the surfaces $F_n'$ must have a common $(n-1)$-fold line $l'$. Two planes in $(x) \sim$ two $F_n'$ in $(x')$ meeting in a residual $C_{2n-1}'$ which meets $l'$ in $2n-2$ points. The line of intersection of the two planes $\sim$ a non-composite $C_n'$ which can meet $l'$ in not more than $n-1$ points. Therefore the residual $C_{n-1}'$ must meet $l'$ in $n-1$ points, i.e. $C_{n-1}'$ consists of $n-1$ lines $l_i'$. The $C_n'$ meets $l'$ in exactly $n-1$ points and each $l_i'$ in one point. Three such surfaces $F_n'$ meet in $n$ points of which $n-1$ must be fixed in order to have a homoloidal web.

Among the surfaces of the web in the $(x)$ space there is a pencil consisting of a plane through $l$ and the fixed $F_{n-1}: l^{n-2}, \Sigma l_i, \Sigma P_i$. Therefore there is a pencil of planes through $l'$ in $(x')$ which corresponds to this pencil of the web of $F_n$, and $l' \sim$ the fixed $F_{n-1}$.

A general line in $(x') \sim$ a $C_n: \Sigma P_i$, hence $P_i \sim$ a plane. Since $C_n$ meets each $l_i$ once, $l_i \sim$ a plane. The plane $\sigma_i$ through $l$ and $P_i \sim$ a plane $\rho_i'$ through $l'$, but since $P_i \sim$ a plane, $P_i \sim \rho_i'$, and the plane $\sigma_i$ apart from $P_i$ and $l \sim$ a curve $s'$ in $\rho_i'$. A general plane of $(x')$ meets $\rho_i'$ in a line $L'$ and the curve $s'$ in one or more points $Q'$. The corresponding $F_n$ meets $\sigma_i$ in one line $L$ through $P_i$. Therefore the curve $s'$ is a line. The line $L'$, except for the point $Q'$, $\sim P_i$ and $Q' \sim L$. The line $s'$ must be a fundamental line $l_i'$ because the points of $s' \sim$ lines in $\sigma_i$ through $P_i$. In a similar manner it is seen that $l_i \sim$ a plane $\sigma_i'$ through $l'$ and $P_i'$. The plane $\rho_i$ through $l$ and $l_i \sim$ the point $P_i'$ in $\sigma_i'$. Therefore the points and lines $P_i$ and $l_i$ are associated in pairs with the lines and points $l_i'$ and $P_i'$ respectively.

3. **The involutorial transformation.** When the two spaces are superimposed for the involutorial case, the fundamental systems must coincide and the planes through $l$ are interchanged in pairs by the involutorial transformation $I_n$. If $x_1 = 0$ and $x_2 = 0$ are the invariant planes of this pencil, the four equations of $I_n$ can be obtained from three homogeneous equations of the following form:

$$(1_1) \qquad x_1' = x_1,$$

$$(1_2) \qquad x_2' = -x_2,$$

$$(1_3) \qquad x_3' = (d + ex_3 + fx_4)/(a + bx_3 + cx_4),$$

* G. Loria, *Sulla classificazione delle trasformazioni di genere zero*, Istituto Lombardo, Rendiconti, (2), vol. 23 (1890), pp. 824–834.

where $a=a(x_1, x_2)$, etc., $(d+ex_3+fx_4)=0$ is any $F_n$ of the web, and $(a+bx_3+c_4)=0$ is the fixed $F_{n-1}$. Since we are dealing with involutorial transformations the inverse of equations (1) have the same form as (1). If in the inverse of $(1_3)$ we replace $x_1'$, $x_2'$ by $x_1$, $-x_2$ we have

$$x_3 = (\bar{d} + \bar{e}x_3' + \bar{f}x_4')/(\bar{a} + \bar{b}x_3' + \bar{c}x_4'), \tag{2}$$

where $\bar{a}=a(x_1, -x_2)$, etc. This equation can be solved for $x_4'$ and thus we get the fourth equation of the involutorial transformation as

$$\begin{aligned} x_4' = \{&[(\bar{e}d + \bar{d}a) - x_3(\bar{b}d + \bar{a}a)] + x_3[(\bar{e}e + \bar{d}b) - x_3(\bar{b}e + \bar{a}b)] \\ &+ x_4[(\bar{e}f + \bar{d}c) - x_3(\bar{b}f + \bar{a}c)]\}/[(a + bx_3 + cx_4)(\bar{c}x_3 - \bar{f})]. \end{aligned} \tag{3}$$

When the conditions that $\bar{c}x_3-\bar{f}$ be a factor of the numerator are satisfied and this factor is removed, we have the $I_n$ with an $(n-1)$-fold line, defined analytically.

4. **The cubic case.** A non-homogeneous coördinate system is useful in the cases when $n=3$ or 4, so we put $x_2/x_1=\lambda$, $x_3/x_1=x$, and $x_4/x_1=y$. When $n=3$ there are only two fundamental points $P_1$, $P_2$; any plane through the line joining them is transformed by $I_3$ into another such plane and the two planes $\rho_i=0$, which are the planes $l$, $l_i$. Among the planes of the pencil on the line $P_1P_2$ there are at least two which are invariant. Let $x=0$ and $y=0$ be two of the invariant planes, and let $\rho_i\equiv\lambda_i-\lambda=0$. The points $P_1$, $P_2$ are then determined by the planes $\sigma_i\equiv\lambda_i+\lambda=0$ and the line $x=y=0$. One surface of the web is $\rho_1\rho_2x=0$, and the equation of the fixed quadric determined by $l$, $2l_i$, $2P_i$ is of the form

$$\sigma_1\sigma_2 + (a_0 + a_1\lambda)x + (b_0 + b_1\lambda)y = 0.$$

We can write the first two equations of $I_3$ as follows:

$$\lambda' = -\lambda, \tag{$4_1$}$$

$$x' = h\rho_1\rho_2x/\{\sigma_1\sigma_2 + (a_0 + a_1\lambda)x + (b_0 + b_1\lambda)y\}. \tag{$4_2$}$$

If we write the inverse of $(4_2)$ and replace $\lambda'$ by $-\lambda$ we have

$$x = h\sigma_1\sigma_2x'/\{\rho_1\rho_2 + (a_0 - a_1\lambda)x' + (b_0 - b_1\lambda)y'\}. \tag{5}$$

When (5) is solved for $y'$, it has the form

$$\begin{aligned} y' = \rho_1\rho_2\{&(h^2 - 1)\sigma_1\sigma_2 - x[a_0(h + 2) - a_1(h - 1)\lambda] \\ &- y(b_0 + b_1\lambda)\}/[\{\sigma_1\sigma_2 + (a_0 + a_1\lambda)x + (b_0 + b_1\lambda)y\}(b_0 - b_1\lambda)]. \end{aligned} \tag{6}$$

Since $y=0$ is an invariant plane the first two terms in the numerator of (6) must vanish, and the coefficient of $y$ must be divisible by $b-b_1\lambda$. This requires that

$$b_1 = 0, \quad h = \pm 1, \quad a_0(h+1) - a_1(h-1)\lambda = 0.$$

The third condition presents two cases, namely

$$h = 1, \quad \text{so that } a_0 = 0,$$

or

$$h = -1, \text{ so that } a_1 = 0.$$

The cubic involutorial transformation may now be written in the form

$$\begin{aligned} \lambda' &= -\lambda, \\ x' &= x\rho_1\rho_2/(\sigma_1\sigma_2 + a_1\lambda x + b_0 y), \\ y' &= -y\rho_1\rho_2/(\sigma_1\sigma_2 + a_1\lambda x + b_0 y), \end{aligned} \tag{7}$$

or

$$\begin{aligned} \lambda' &= -\lambda, \\ x' &= -x\rho_1\rho_2/(\sigma_1\sigma_2 + a_0 x + b_0 y), \\ y' &= -y\rho_1\rho_2/(\sigma_1\sigma_2 + a_0 x + b_0 y). \end{aligned} \tag{8}$$

In the first case when $h=1$, the pencil of planes through $P_1$, $P_2$ is invariant. In the second case when $h=-1$, each plane of the pencil is invariant.

5. **The quartic case.** There are three fundamental points; one of the surfaces of the web consists of the plane $P_1P_2P_3$ and of the three planes $l$, $l_i$. We can take the points $P_1$, $P_2$ as in $I_3$ and the plane $P_1P_2P_3$ as $x=0$. The lines $l_i$ in the planes $\rho_i \equiv \lambda_i - \lambda = 0$ and the points $P_i$ lie in the planes $\sigma_i \equiv \lambda_i + \lambda = 0$. The equation of the fixed $F_3 : l^2, 3l_i, 3P_i$ may be written in the form

$$\sigma_1\sigma_2(1-\lambda) + (a_0 + a_1\lambda + a_2\lambda^2)x + (b_0 + b_1\lambda + b_2\lambda^2)y = 0.$$

The first two equations of the $I_4$ are now given by

$$(9_1) \quad \lambda' = -\lambda,$$

$$(9_2) \quad x' = h\rho_1\rho_2\rho_3 x/[\sigma_1\sigma_2(1-\lambda) + (a_0 + a_1\lambda + a_2\lambda^2)x + (b_0 + b_1\lambda + b_2\lambda^2)y].$$

The inverse of $(9_2)$ with $\lambda'$ replaced by $-\lambda$ is

$$x = h\sigma_1\sigma_2\sigma_3 x'/[\rho_1\rho_2(1+\lambda) + (a_0 - a_1\lambda + a_2\lambda^2)x' + (b_0 - b_1\lambda + b_2\lambda^2)y']. \tag{10}$$

If (10) is solved for $y'$ we have

$$\begin{aligned} y' = \rho_1\rho_2\{&\sigma_1\sigma_2(h^2\rho_3\sigma_3 - 1 + \lambda^2) - x[h\rho_3(a_0 - a_1\lambda + a_2\lambda^2) \\ &+ (1+\lambda)(a_0 + a_1\lambda + a_2\lambda^2)] - y(1+\lambda)(b_0 + b_1\lambda + b_2\lambda^2)\}/\{[b_0 \\ &- b_1\lambda + b_2\lambda^2][\sigma_1\sigma_2(1-\lambda) + (a_0 + a_1\lambda + a_2\lambda^2)x + (b_0 + b_1\lambda + b_2\lambda^2)y]\}. \end{aligned} \tag{11}$$

The expressions

$$\sigma_1\sigma_2(h^2\rho_3\sigma_3 - 1 + \lambda^2), \tag{12}$$

$$h\rho_3(a_0 - a_1\lambda + a_2\lambda^2) + (1 + \lambda)(a_0 + a_1\lambda + a_2\lambda^2), \tag{13}$$

$$(1 + \lambda)(b_0 + b_1\lambda + b_2\lambda^2) \tag{14}$$

must therefore contain the factor $b_0 - b_1\lambda + b_2\lambda^2$. From (14) we find that $b_1 = 0$ and if we use this in (12) we have the condition

$$(1 - h^2)/(\lambda_3^2 h^2 - 1) = b_2/b_0. \tag{15}$$

From (13) we get the conditions

$$[a_2(1 + h\lambda_3) + a_1(1 + h)]/[a_0(1 + h\lambda_3)] = b_2/b_0, \tag{16}$$

$$(1 - h)a_2/[a_1(1 - h\lambda_3) + a_0(1 - h)] = b_2/b_0. \tag{17}$$

These last two conditions may be rewritten as

$$(h\lambda_3 + 1)(a_0b_2 - a_2b_0) = a_1b_0(1 + h), \tag{16'}$$

$$(1 - h)(a_0b_2 - a_2b_0) = a_1b_2(h\lambda_3 - 1). \tag{17'}$$

If we divide (17′) by (16′) we get condition (15) over again so that (15) is included in (16) and (17). We can solve (16) and (17) for $h$ and obtain

$$h = (a_0b_2 - a_2b_0 + a_1b_2)/(a_0b_2 - a_2b_0 + \lambda_3a_1b_2), \tag{18}$$

$$h = (a_1b_1 - a_0b_2 + a_2b_0)/(\lambda_3a_0b_2 - \lambda_3a_2b_0 - a_1b_0); \tag{19}$$

if we equate these values of $h$ we get

$$(\lambda_3 + 1)(a_0b_2 - a_2b_0)^2 + 2a_1(a_0b_2 - a_2b_0)(b_2\lambda_3 - b_0) - a_1^2b_0b_2(\lambda_3 + 1) = 0. \tag{20}$$

The quartic involutorial transformation is therefore determined by the equations

(21)

$$\lambda = -\lambda,$$

$$x' = \frac{hx\rho_1\rho_2\rho_3}{\sigma_1\sigma_2(1 - \lambda) + (a_0 + a_1\lambda + a_2\lambda^2)x + (b_0 + b_2\lambda^2)y},$$

$$y' = \frac{\rho_1\rho_2\{\sigma_1\sigma_2(1 - h^2) - x[a_2(1 + h\lambda_3) + a_1(1 + h) + a_2\lambda(1 - h) - b_2y(1 + \lambda)\}}{b_2[\sigma_1\sigma_2(1 - \lambda) + (a_0 + a_1\lambda + a_2\lambda^2)x + (b_0 + b_2\lambda^2)y]}.$$

In these equations $h$ is defined by (18) or (19) and the coefficients $a_i$ and $b_i$ are subject to condition (20).

6. **The quintic case.** There is a net of quadrics through $l$ and the $4P_i$ which is invariant under the $I_5$. We can use the vertices of the tetrahedron of reference for the $4P_i$ and take

$$X_1 \equiv d_1x_1 + d_2x_2 + d_3x_3 + d_4x_4 = 0,$$
$$X_2 \equiv x_1 + x_2 + x_3 + x_4 = 0$$

as the invariant planes through $l$, so that

$$x_1' = X_1F_4, \qquad x_2' = -X_2F_4, \tag{22}$$

where $F_4$ is determined by $l^3$, $4l_i$, $4P_i$. The planes $l$, $P_i$ are given by $\sigma_i \equiv X_1 - d_iX_2 = 0$ and the planes $l$, $l_i$ by $\rho_i \equiv X_1 + d_iX_2 = 0$. The net of quadrics has the form

$$k_1\sigma_1x_1 + k_2\sigma_2x_2 + k_3\sigma_3x_3 = 0$$

and from (22) we have the identity

$$\sigma_1x_1 + \sigma_2x_2 + \sigma_3x_3 + \sigma_4x_4 = 0.$$

The quadrics of the net are interchanged in pairs involutorially by $I_5$, so that the involutorial transformation can be defined by

$$\begin{aligned}
(a_1\sigma_1'x_1' + a_2\sigma_2'x_2' + a_3\sigma_3'x_3') &= (a_1\sigma_1x_1 + a_2\sigma_2x_2 + a_3\sigma_3x_3)\rho_2\rho_3\rho_4F_4,\\
(b_1\sigma_1'x_1' + b_2\sigma_2'x_2' + b_3\sigma_3'x_3') &= (b_1\sigma_1x_1 + b_2\sigma_2x_2 + b_3\sigma_3x_3)\rho_1\rho_3\rho_4F_4,\\
(c_1\sigma_1'x_1' + c_2\sigma_2'x_2' + c_3\sigma_3'x_3') &= -(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)\rho_1\rho_2\rho_4F_4,
\end{aligned} \tag{23}$$

and the identity

$$\sigma_1'x_1' + \sigma_2'x_2' + \sigma_3'x_3' + \sigma_4'x_4' = 0. \tag{24}$$

If we solve (23) for $x_i'$ replacing $\sigma_i'$ by $\rho_i$ and use (24) to obtain $x_4'$, we have the $I_5$ expressed by

$$\begin{aligned}
x_1' &= [\sigma_1x_1\Delta - 2C_1(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)]\rho_2\rho_3\rho_4,\\
x_2' &= [\sigma_2x_2\Delta - 2C_2(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)]\rho_1\rho_3\rho_4,\\
x_3' &= [\sigma_3x_3\Delta - 2C_3(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)]\rho_1\rho_2\rho_4,\\
x_4' &= [\sigma_4x_4\Delta + 2(C_1 + C_2 + C_3)(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)]\rho_1\rho_2\rho_3,
\end{aligned} \tag{25}$$

where

$$\Delta = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix},$$

and $C_i$ is the cofactor of $c_i$ in $\Delta$.

7. **The involutorial transformations $I_6$, $I_7$, and $I_n$.** There is a net of $F_3{:}l^2$, $5P_i$, $l_5$ which is transformed into a net of $F_2{:}l$, $4P_i (i<5)$ by $I_6$. Among the cubics of the net there is the pencil of $F_2{:}l$, $5P_i$ with the fixed component

$\rho_5=0$, which is invariant under $I_6$. Hence using the same coordinate system as in the quintic case we can determine the $I_6$ by

$$\begin{aligned}
a_1\sigma_1' x_1' + a_2\sigma_2' x_2' + a_3\sigma_3' x_3' &= (a_1\sigma_1 x_1 + a_2\sigma_2 x_2 + a_3\sigma_3 x_3)\rho_1\rho_2\rho_3\rho_4\rho_5 F_5,\\
b_1\sigma_1' x_1' + b_2\sigma_2' x_2' + b_3\sigma_3' x_3' &= -(b_1\sigma_1 x_1 + b_2\sigma_2 x_2 + b_3\sigma_3 x_3)\rho_1\rho_2\rho_3\rho_4\rho_5 F_5,\\
c_1\sigma_1' x_1' + c_2\sigma_2' x_2' + c_3\sigma_3' x_3' &= F_3\rho_1\rho_2\rho_3\rho_4 F_5,\\
d_1x_1' + d_2x_2' + d_3x_3' + d_4x_4' &= (d_1x_1 + d_2x_2 + d_3x_3 + d_4x_4)F_5,\\
x_1' + x_2' + x_3' + x_4' &= -(x_1 + x_2 + x_3 + x_4)F_5,
\end{aligned}$$

where $F_5$ is the fixed quintic surface. The $a_i$ and $b_i$ are restricted since these quadrics must contain $P_5$. The cubic $F_3$ is of the form

$$F_3 \equiv (g_1X_1^2 + g_2X_1X_2 + g_3X_2^2)(g_4x_1 + g_5x_2 + g_6x_3) + \rho_5(g_7\sigma_1x_1 + g_8\sigma_2x_2 + g_9\sigma_3x_3) = 0,$$

where $g_4x_1+g_5x_2+g_6x_3=0$ is the plane through $l_5$, $P_4$; and $g_7\sigma_1x_1+g_8\sigma_2x_2+g_9\sigma_3x_3=0$ is a quadric of the pencil $l$, $5P_i$. The $a_i$, $b_i$, $c_i$, $d_i$, $g_i$ must satisfy the conditions necessary in order that $F_3$ may be transformed by $I_6$ into

$$(c_1\sigma_1x_1 + c_2\sigma_2x_2 + c_3\sigma_3x_3)\rho_1\rho_2\rho_3\rho_4\rho_5\sigma_5F_6^2.$$

When $n=7$ there is a net of quartic surfaces $F_4:l^3$, $6P_i$, $l_5$, $l_6$ which correspond to a net of quadrics $F_2:l$, $4P_i(i<5)$. Among the surfaces of the net there is a pencil of $F_3:l^2$, $6P_i$, $l_5$ with $\rho_6=0$ as a fixed component, which is transformed into the pencil of $F_2:l$, $5P_i(i<6)$. Among the surfaces of the pencil there is the cubic consisting of the plane $\rho_5=0$ and the quadric $F_2:l$, $6P_i$ which is invariant under $I_7$. The equations which determine the $I_7$ are therefore of the following form:

$$\begin{aligned}
a_1\sigma_1' x_1' + a_2\sigma_2' x_2' + a_3\sigma_3' x_3' &= (a_1\sigma_1 x_1 + a_2\sigma_2 x_2 + a_3\sigma_3 x_3)\rho_1\rho_2\rho_3\rho_4\rho_5\rho_6 F_6,\\
b_1\sigma_1' x_1' + b_2\sigma_2' x_2' + b_3\sigma_3' x_3' &= F_3\rho_1\rho_2\rho_3\rho_4\rho_5 F_6,\\
c_1\sigma_1' x_1' + c_2\sigma_2' x_2' + c_3\sigma_3' x_3' &= F_4\rho_1\rho_2\rho_3\rho_4 F_6,\\
d_1x_1' + d_2x_2' + d_3x_3' + d_4x_4' &= (d_1x_1 + d_2x_2 + d_3x_3 + d_4x_4)F_6,\\
x_1' + x_2' + x_3' + x_4' &= -(x_1 + x_2 + x_3 + x_4)F_6.
\end{aligned}$$

The forms obtained for $n=5$, 6, 7 can be generalized as follows:

A. If $n=3m-1$ there is a net of $F_m:l^{m-1}$, $\Sigma P_i$.

B. If $n=3m$ there is a net of $F_{m+1}:l^m$, $\Sigma P_i$, $l_{3m-1}$ containing the pencil of $F_m:l^{m-1}$, $\Sigma P_i$ with a fixed component the plane $l$, $l_{3m-1}$.

C. If $n=3m+1$ there is a net of $F_{m+2}:l^{m+1}$, $\Sigma P_i$, $l_{3m-1}$, $l_{3m}$ containing a pencil of $F_{m+1}:l^m$, $\Sigma P_i$, $l_{3m-1}$ with the fixed component $l$, $l_{3m}$. One of the surfaces of the pencil is the $F_m:l^{m-1}$, $\Sigma P_i$ with the fixed component $l$, $l_{3m-1}$.

In each case the net is transformed into a net of $F_m: l^{m-1}$, $\Sigma P_i (i < 3m-1)$ by $I_n$, hence $I_n$ can be defined by means of the nets.

8. **The mapping of the involutorial transformation $I_n$.** The expressions $2x_1x_1'$, $-2x_2x_2'$, $x_1x_3' + x_3x_1'$, $x_2x_3' + x_3x_2'$ are invariant under $I_n$. Let us consider the correspondence between the $(x)$ space and a $(y)$ space where the values of $x_i'$ above are those defined in §3, (1). The correspondence has the form

$$
\begin{aligned}
y_1 &= 2x_1^2(a + bx_3 + cx_4), \\
y_2 &= 2x_2^2(a + bx_3 + cx_4), \\
y_3 &= x_1[(d + ex_3 + fx_4) + x_3(a + bx_3 + cx_4)], \\
y_4 &= x_2[(d + ex_3 + fx_4) - x_3(a + bx_3 + cx_4)].
\end{aligned} \tag{26}
$$

These equations can be solved for $x_i$ as follows:

$$
\begin{aligned}
x_2/x_1 &= \pm (y_2/y_1)^{1/2}, \\
x_3 &= -x_1\overline{U}/(x_2y_1), \\
x_4 &= [dx_2^2y_1^2 - ax_1x_2y_1U - ex_1x_2y_1\overline{U} + bx_1^2U\overline{U}]/[cx_1x_2y_1U - fx_2^2y_1^2],
\end{aligned} \tag{27}
$$

where $U = x_1y_4 + x_2y_3$, $\overline{U} = x_1y_4 - x_2y_3$. Hence equations (26) define a (1, 2) correspondence.

If we rewrite the equations of the correspondence in terms of $x_i'$ and replace $x_1'$, $x_2'$ by $x_1$, $-x_2$, we have

$$
\begin{aligned}
y_1 &= 2x_1^2(\bar{a} + \bar{b}x_3' + \bar{c}x_4'), \\
y_2 &= 2x_2^2(\bar{a} + \bar{b}x_3' + \bar{c}x_4'), \\
y_3 &= x_1[(\bar{d} + \bar{e}x_3' + \bar{f}x_4') + x_3'(\bar{a} + \bar{b}x_3' + \bar{c}x_4')], \\
y_4 &= -x_2[(\bar{d} + \bar{e}x_3' + \bar{f}x_4') - x_3'(\bar{a} + \bar{b}x_3' + \bar{c}x_4')],
\end{aligned} \tag{28}
$$

where $\bar{a} = a(x_1, -x_2)$, etc. If we equate the values of $y_i$ given in (26) and (28) we have

$$
\begin{aligned}
a + bx_3 + cx_4 &= \bar{a} + \bar{b}x_3' + \bar{c}x_4', \\
(d + ex_3 + fx_4) + x_3(a + bx_3 + cx_4) &= (\bar{d} + \bar{e}x_3' + \bar{f}x_4') + x_3'(\bar{a} + \bar{b}x_3' + \bar{c}x_4'), \\
(d + ex_3 + fx_4) - x_3(a + bx_3 + cx_4) &= -(\bar{d} + \bar{e}x_3' + \bar{f}x_4') + x_3'(\bar{a} + \bar{b}x_3' + \bar{c}x_4').
\end{aligned} \tag{29}
$$

From (29) we get

$$
\begin{aligned}
x_3' &= (d + ex_3 + fx_4)/(a + bx_3 + cx_4), \\
x_3 &= (\bar{d} + \bar{e}x_3' + \bar{f}x_4')/(\bar{a} + \bar{b}x_3' + \bar{c}x_4'),
\end{aligned} \tag{30}
$$

but these are precisely the equations §3, (1), (2) by which the involutorial transformation $I_n$ was defined.

Hence we see that a (1, 2) correspondence of the type given by equations (26) leads in general to a special type of involutorial transformation of order $2n-1$ with a $(2n-3)$-fold line. If however conditions are imposed that $\bar{c}x_3-\bar{f}$ be a factor of the numerator of the expression for $x_4'$, then $\bar{c}x_3-\bar{f}$ is a factor of $x_1'$, $x_2'$, $x_3'$, and $x_4'$ and we have an $I_n$ with an $(n-1)$-fold line. We have therefore proved the

THEOREM. *An involutorial transformation in $S_3$ of order $n$ with an $(n-1)$-fold line is rational.*

9. **Image of a general line in $(y)$.** A line $y_3=Ay_1+By_2$, $y_4=Cy_1+Dy_2$ in the $(y)$ space is transformed by the correspondence into the $C_{n+4}$ given parametrically by the equations

$$
\begin{aligned}
X_1 &= x_1^2 x_2(cV - fx_1x_2),\\
X_2 &= x_1x_2^2(cV - fx_1x_2),\\
X_3 &= -\overline{V}(cV - fx_1x_2),\\
X_4 &= dx_1^2x_2^2 - x_1x_2(aV + e\overline{V}) + bV\overline{V},
\end{aligned}
\tag{31}
$$

where

$$
\begin{aligned}
V &= x_1(Cx_1^2 + Dx_2^2) + x_2(Ax_1^2 + Bx_2^2),\\
\overline{V} &= x_1(Cx_1^2 + Dx_2^2) - x_2(Ax_1^2 + Bx_2^2).
\end{aligned}
$$

In the case of the $I_3$ the curve is a $C_7$ of the form

$$
\begin{aligned}
X_1 &= b_0x_1^3x_2V,\\
X_2 &= b_0x_1^2x_2^2V,\\
X_3 &= -b_0x_1V\overline{V},\\
X_4 &= x_2[a_1V\overline{V} - x_1(\sigma_1\sigma_2V + \rho_1\rho_2\overline{V})].
\end{aligned}
\tag{32}
$$

If we put $V=0$ in (32), there are three values of the parameter $x_2/x_1$ all giving the point (0, 0, 0, 1). If we put $\overline{V}=0$, we get three distinct points in the invariant plane $x_3=0$. When $x_1=0$, we again have the point (0, 0, 0, 1) and furthermore the $C_7$ is tangent to $l$ at that point with $x_1=0$ as the osculating plane. When we put $x_2=0$, we have the point (0, 0, 1, 0). Hence the $C_7$ has a fourfold point (0, 0, 0, 1) at which it is also tangent to $l$ and passes through the point (0, 0, 1, 0).

In the general case there are $n+1$ values of $x_2/x_1$ due to the vanishing of $(cV-fx_1x_2)$ which give the point (0, 0, 0, 1). When $x_1=0$ or $x_2=0$, we get

two definite points on $l$ at which the $C_{n+4}$ is tangent to the planes $x_1=0$, $x_2=0$ respectively. A plane of the pencil $y_2=k^2y_1$ has for images the two planes $x_2=\pm kx_1$ which meet the $C_{n+4}$ in the two images of the point in which $y_2=k^2y_1$ meets the line of which the $C_{n+4}$ is the image.

10. **Image of a line in $(y)$ which meets $l'$.** Any line in $(y)$ meeting $l'$ may be defined by

$$y_2 = k^2y_1, \quad Ay_1 + By_2 + Cy_3 + Dy_4 = 0.$$

The image in $(x)$ of such a line is a pair of conics each belonging to a net in the planes $x_2=\pm kx_1$. In the plane $x_2=kx_1$ the net has the form

$$2x_1(a_{11}x_1 + b_{11}x_3 + c_{11}x_4)(A + k^2B) + x_1(d_{11}x_1 + e_{11}x_3 + f_{11}x_4)(C + kD) + x_3(a_{11}x_1 + b_{11}x_3 + c_{11}x_4)(C - kD) = 0,$$

where $a_{11}=a(1, k)$, etc. The conics of the net pass through the fixed points

$$x_1 = x_2 = b_{11}x_3 + c_{11}x_4 = 0,$$

$$x_1 = x_2 = x_3 = 0,$$

$$x_2 - kx_1 = a_{11}x_1 + b_{11}x_3 + c_{11}x_4 = d_{11}x_1 + e_{11}x_3 + f_{11}x_4 = 0.$$

Two lines in $y_2=k^2y_1$ have for images a pair of conics of each net; the point of intersection of the two lines corresponds to the two free intersections of the two pairs of conics.

In the case of the invariant plane $y_2=0$, the lines in $y_2=0$ correspond to a pencil of conics in the plane $x_2=0$ given by

$$2Ax_1(a_{10}x_1+b_{10}x_3+c_{10}x_4)+C[x_1(d_{10}x_1+e_{10}x_3+f_{10}x_4)+x_3(a_{10}x_1+b_{10}x_3+c_{10}x_4)]=0$$

where $a_{10}=a(1, 0)$, etc. The pencil of conics has the three fixed points

$$x_1 = x_2 = b_{10}x_3 + c_{10}x_4 = 0,$$

$$x_1 = x_2 = x_3 = 0,$$

$$x_2 = a_{10}x_1 + b_{10}x_3 + c_{10}x_4 = d_{10}x_1 + e_{10}x_3 + f_{10}x_4 = 0.$$

The variable point of intersection of the net of conics is in this case replaced by the direction of the tangent to

$$x_1(d_{10}x_1 + e_{10}x_3 + f_{10}x_4) + x_3(a_{10}x_1 + b_{10}x_3 + c_{10}x_4) = 0$$

at the point $x_1=x_2=b_{10}x_3+c_{10}x_4=0$. Hence this point is an invariant point the image of which in $(y)$ is $y_2=0$. Similarly the plane $y_1=0$ is the image of the invariant point $x_1=x_2=b_{01}x_3+c_{01}x_4=0$. The surface of branch points in the $(y)$ space consists of the two planes $y_1=0$, $y_2=0$, and the corresponding surface of coincidences in $(x)$ reduces to the two invariant points.

## ADDENDUM

In a recently published article* Snyder discusses involutorial birational transformations contained multiply in a linear line complex and suggests that they are probably irrational. The transformation he considers is of order $2k$ with a $(2k-1)$-fold line $x_3=x_4=0$, and $2k-1$ fundamental points lying on the line $x_1=x_2=0$, and so is a special case of the involutorial transformations studied in this paper. The equations of the $I_{2k}$ are given as

$$\begin{aligned} x_1' &= (x_3^{2k-1} + x_4^{2k-1})x_1, \\ x_2' &= (x_3^{2k-1} + x_4^{2k-1})x_2, \\ x_3' &= (x_4^{2k-1} - x_3^{2k-1})x_3, \\ x_4' &= (x_3^{2k-1} - x_4^{2k-1})x_4. \end{aligned}$$

This involutorial transformation may be mapped, as in the general case, on ordinary space by the (1, 2) correspondence given by the equations

$$\begin{aligned} y_1 &= x_1x_4x_3^{2k-1}, \\ y_2 &= x_2x_4x_3^{2k-1}, \\ y_3 &= x_3^2(x_3^{2\,-1} - x_4^{2k-1}), \\ y_4 &= x_4^2(x_3^{2\,-1} - x_4^{2k-1}), \end{aligned}$$

and is therefore rational.

---

* V. Snyder, *The simplest involutorial transformation contained multiply in a line complex*, Bulletin of the American Mathematical Society, vol. 36 (1930), pp. 89–93.

CORNELL UNIVERSITY,
ITHACA, N. Y.

# THE LIMIT OF TRANSITIVITY OF A SUBSTITUTION GROUP*

BY

MARIE J. WEISS†

1. Jordan published his final contribution to the problem of the limit of transitivity of a substitution group which does not contain the alternating group in Liouville's Journal of 1895.‡ Here he gave the interesting inequalities between $n$, the degree, and $t$, the multiplicity of transitivity, of a $t$-ply transitive group, stated in the following theorem:

*Let $n$ be the degree of a $t$-ply transitive group $G$ of class $>3$. Then if $t \geqq 8$,*

*$n-t \geqq 2^{\alpha}$, $\alpha$ an integer $\geqq k-3-\log k/\log 2$, and $k$ an integer such that $5 \leqq k \leqq t$; or*

*$n-t \geqq t!/\{\delta!(t-\delta)!\}$, $\delta$ being the greatest integer less than the quantity $t-(t-k+1)\log 2/(k+\log 2)$.*

This paper followed Bochert's study of the problem in the Mathematische Annalen of 1887 and 1889. Among the results of this study was the inequality

$$\log n \geqq a(t \log t)^{1/2},$$

where the constant $a$ may be taken to be $(1/8)\{(\log 2)/8\}^{1/2}$, if $t \geqq 8$.§

In recent years Miller has given still another relation between $n$ and $t$, namely, $n \geqq (4/25)(t+2)^2$.‖

In this paper inequalities similar to those of Jordan will be established. To obtain these it was necessary to study separately the $t$-ply transitive group whose subgroup that fixes $t$ letters is of order a power of two and the $t$-ply transitive group whose subgroup that fixes $t$ letters has its order divisible by an odd prime. The methods used in the study of the former group, though suggested by Jordan's 1895 paper, are new, while those used in the investigation of the latter group follow to a certain extent those of Jordan. The following theorems give the chief results of this study:

*Let the subgroup that fixes $t$ letters of a $t$-ply transitive group of degree $n$ and class $>3$ be of order $2^m$. Then if $t \geqq 8$,*

$$(n-t)/2 \geqq t!/\{\beta!(t-\beta)!\},$$

* Presented to the Society, August 29, 1929; received by the editors in August, 1929.

† National Research Fellow.

‡ C. Jordan, Journal de Mathématiques, (5), vol. 1 (1895), pp. 35–60.

§ A. Bochert, Mathematische Annalen, vol. 33 (1889), p. 577.

‖ G. A. Miller, Proceedings of the National Academy of Sciences, vol. 2 (1916), p. 61.

*where $\beta$ is an integer chosen such that $t/2<\beta\leqq t-3$, or chosen to be one unit less than the least value of $\beta$ which satisfies the inequality*

$$[(\beta+1)/2]!\geqq 2t!/\{\beta!(t-\beta)!\};$$

*or*

$$n-t\geqq[(t+1)/2]!;$$

*or*

$$n-t\geqq t!/[t/2]!^2.$$

*The symbol $[s]$ denotes the integral part of $s$.*

*Let the order of the subgroup that fixes $t$ letters of a $t$-ply transitive group of degree $n$ and class $>3$ be divisible by an odd prime. Then if $t\geqq 8$,*

$$(n-2t+1)/2\geqq 2^{\alpha},\ \alpha \text{ an integer} > 3 \text{ and } \geqq t-3-\log t/\log 2;$$

*or*

$$(n-2t+1)/2\geqq 2^{\gamma}t,\ \gamma \text{ an integer} > 3 \text{ and } \geqq k-3-\log k/\log 2,$$

*where $k$ is an integer such that $5\leqq k\leqq t-1$; or*

$$(n-2t+1)/2\geqq t!/\{\delta!(t-\delta)!\},$$

*where $\delta$ is the greatest integer less than the quantity $t-(t-k+1)\log 2/(k+\log 2)$.*

It may be of interest to see how Jordan's, Miller's, and the author's results compare for a few values of $t$. In the following table the minimum value of $n$ has been calculated for a given $t$ from the above formulas:

| $t$ | 8 | 16 | 25 | 50 | 100 |
|---|---|---|---|---|---|
| Jordan | 16 | 32 | 50 | 1,275 | 161,800 |
| Miller | 16 | 52 | 117 | 433 | 1,665 |
| Author | 31 | 271 | 849 | 6,499 | 409,799 |

Bochert's inequality is of no interest for these small values of $t$.

2. The first step in the study of the $t$-ply transitive group $G$ is the proof of the following lemma:

LEMMA. *If $t>5$, the subgroup that fixes $t$ letters of a $t$-ply transitive group fixes exactly $t$, $t+1$, or $t+2$ letters.*

Let $G_x$ be the subgroup that fixes $x$ letters of a $t$-ply transitive group $G$. Suppose that $G_t$ fixes the $t+r-1$ letters $b_1, b_2, \cdots, b_t, a_1, a_2, \cdots, a_{r-1}$, $r>1$. Now consider the largest subgroup of $G_{t-1}$ in which $G_t$ is invariant. Its order is $rg_t$, where $g_t$ is the order of $G_t$, and it has a regular constituent of degree $r$ on the letters $b_1, a_1, a_2, \cdots, a_{r-1}$. Since all the subgroups similar to $G_t$ found in $G_{t-1}$ are conjugate, all the subgroups similar to $G_t$ in any of the preceding groups $G_{t-i}$, $i=2, 3, \cdots, t$, $G_0=G$, form one complete set of conjugates. Then the largest subgroup of $G_{t-2}$ in which $G_t$ is invariant has a doubly transitive constituent of degree $r+1$ on the letters $b_2, b_1, a_1, a_2, \cdots,$

$a_{r-1}$.* Finally, the largest subgroup of $G$ in which $G_t$ is invariant has a $t$-ply transitive constituent of degree $r+t-1$ and order $r(r+1)\cdots(r+t-1)$ on the letters that $G_t$ fixes. Now Jordan has shown that such a non-alternating $t$-ply transitive group does not exist unless $t \leqq 3$, or $t=4$ and $r=8$, or $t=5$ and $r=8$.† Hence if $t>5$, the $t$-ply transitive constituent is alternating or symmetric and $r=2$, or 3.

In the remaining part of the paper it will be assumed that $t>5$. The study of the problem will be divided into two parts. In the first part the special case $G_t$ of order $2^m$ will be studied, while in the second part the more general case $G_t$ of order $p^\alpha q$, $p$ an odd prime, will be investigated. It will be found that the former case leads to relations between $n$ and $t$ which give more favorable results than those derived from the more general case. Thus when $t$ is sufficiently large, the former relations may be neglected.

## $G_t$ OF ORDER $2^m$

3. Suppose that $G_t$ is of order $2^m$ and that it contains a transitive subgroup. Let $M$ of degree $2^\gamma$ be its largest transitive subgroup. Now $G_t$ cannot have two transitive subgroups on different sets of letters, for $2^\gamma \geqq u$, the class of $G$, and $u \geqq (n-1)/2$.‡ Then let $G_{t-1}$ be of degree $2^\gamma+q$. From the theory of primitive groups with transitive subgroups of lower degree,§ it is known that $q$ divides $2^\gamma$, and that $G_{t-1}$ is imprimitive with systems of imprimitivity of $q$ letters. Hence $G_t$ fixes exactly $t+1$ letters, for $q$ would be odd in case $G_t$ fixed $t$ or $t+2$ letters, the only other possibilities by the above lemma. Now let the letters displaced by $G_{t-1}$ but fixed by $M$ be $a_1'$, $a_2'$, $\cdots$, $a_q'$; the letters displaced by $G_{t-1}$ but fixed by $G_t$ be $a_1'$ and $a_2'$; the letters introduced by $G_{t-2}$, $G_{t-3}$, $\cdots$, $G$, respectively, be $b$, $b_1$, $\cdots$, $b_{t-2}$. Then the letters $a_1'$, $a_2'$, $\cdots$, $a_q'$ form one system of imprimitivity of $G_{t-1}$, and let $a_1$, $a_2$, $\cdots$, $a_q$ be the letters of another system of imprimitivity of $G_{t-1}$.

Now consider the largest group $I$ of $G$ in which $G_t$ is invariant. It transforms $M$ into itself, for $M$ is the largest transitive subgroup of $G_t$. If $I$ did not transform $M$ into itself, it would transform $M$ into a conjugate $M'$ which, as has been seen in the above paragraph, is not entirely free from

* W. A. Manning, Bulletin of the American Mathematical Society, vol. 13 (1906), p. 20.

† C. Jordan, Journal de Mathématiques, (2), vol. 17 (1872), p. 351.

‡ W. A. Manning, these Transactions, vol. 31 (1929), p. 648.

§ C. Jordan, Journal de Mathématiques, (2), vol. 16 (1871), pp. 383–408.

Marggraff, Dissertation, *Ueber primitive Gruppen mit transitiven Untergruppen geringeren Grades*, Giessen, 1889.

W. A. Manning, these Transactions, vol. 7 (1906), pp. 499–508.

W. A. Manning, *Primitive Groups*, 1921, Chap. VI. The reader is referred to this last reference for a brief yet complete account of the theory.

the letters of $M$. The transitive groups $M$ and $M'$ then would generate a transitive subgroup of degree or order greater than $M$.

Further $I$ has a symmetric constituent on the letters $a_1'$, $a_2'$, $b$, $b_1$, $\cdots$, $b_{t-2}$. Hence it contains the substitution

$$T = (bb_1b_2)(a_1')(a_2')(b_3)(b_4)\cdots(b_{t-2})\cdots.$$

It may be assumed that $T$ fixes $a_1$, for if $T$ displaces the letter $a_1$, the product of $T$ and a properly chosen substitution from $M$ gives a substitution fixing $a_1$. Since the set of letters $a_1'$, $a_2'$, $\cdots$, $a_q'$ and the set of letters $a_1$, $a_2$, $\cdots$, $a_q$ are the letters of two systems of imprimitivity of $G_{t-1}$, $T$ permutes the letters of each of these sets among themselves. Proper powers of $T$ will reduce it to a substitution of order three. It will be shown that this substitution occurs in a conjugate of $G_t$. Now it cannot displace all the letters $a_2$, $a_3$, $\cdots$, $a_q$ and all the letters $a_3'$, $a_4'$, $\cdots$, $a_q'$, for then $q-2\equiv 0$, mod 3, and $q-1\equiv 0$, mod 3. Thus this substitution fixes another letter $a_x$. Hence it occurs in the $G_t$ which fixes $b_3$, $b_4$, $\cdots$, $b_{t-2}$, $a_1'$, $a_2'$, $a_1$, and $a_x$. The following theorem has then been proved:

THEOREM 1. *If $t>5$, and if the subgroup which fixes $t$ letters of a $t$-ply transitive group $G$ is of order $2^m$, $G$ contains no transitive subgroup of degree $<n-t+1$.*

4. Let $I'$ be that subgroup of $G$, in which $G_t$ is invariant, which has an alternating constituent on the $t$ letters $b_1$, $b_2$, $\cdots$, $b_t$ that $G_t$ fixes. Let $a$'s be the letters that $G_t$ displaces. Further assume that $I'$ permutes among themselves the letters of each transitive constituent of $G_t$. Then each substitution $(b_1b_2b_3)(b_4)(b_5)\cdots(b_t)(a\cdots)\cdots$ of order three of $I'$ fixes at least one letter of each transitive constituent of $G_t$, for each such transitive constituent is of degree a power of two. Hence $G_t$ has exactly two transitive constituents, for if it had three or more, these substitutions would occur in the $G_t$ which fixed the three or more $a$'s of the transitive constituents fixed and the letters $b_4$, $b_5$, $\cdots$, $b_t$. The case when $G_t$ has only one transitive constituent falls under the previous discussion (§3). Moreover $G_t$ fixes exactly $t(>5)$ letters, for if it fixed more, the substitution $(b_1b_2b_3)(b_4)(b_5)\cdots(b_t)(a_u)(a_v)\cdots$ of order three would occur in the $G_t$ which fixed the letters $b_4$, $b_5$, $\cdots$, $b_t$, $a_u$, $a_v$, $x$, where $a_u$ and $a_v$ are the letters of $G_t$ that such a substitution fixes and $x$ is the additional letter fixed by $G_t$.*

If $G_{t-1}$ is imprimitive it has systems of imprimitivity of degree $2^\gamma+1$,

* If $t=5$, $G_t$ (see lemma, §2) may fix 5, 6, 7, or 12 letters. However in the last three cases the substitution $(b_1b_2b_3)(b_4)\cdots(b_t)(a_u)(a_v)\cdots$ of order three again fixes more than $t$ letters. Thus $G_t$ also fixes exactly $t$ letters if $t=5$.

where $2^\gamma$ is the degree of one of the transitive constituents of $G_t$. Let the other transitive constituent of $G_t$ be of degree $2^\delta$. Then $2^\gamma+1$ divides $2^\gamma+1+2^\delta$, the degree of $G_{t-1}$. This division is obviously impossible, and hence $G_{t-1}$ is primitive.

Now consider the subgroup $F$ of $G_t$ which fixes one letter of one of the transitive constituents of $G_t$. If $F$ reduces to the identity, $G_t$ has a regular constituent simply or multiply isomorphic to the second transitive constituent. If there is a multiple isomorphism between the constituents, let $F$ be the subgroup which fixes one letter from the constituent of smaller order. If the transitive constituents are in simple isomorphism, each constituent is regular or one is not. The former case will be considered later. In the latter case take $F$ to be the subgroup which fixes one letter of the non-regular constituent. Thus, unless $G_t$ is a simple isomorphism between two regular constituents, a subgroup $F$ of $G_t$ may be found which does not reduce to the identity.

Now let $X$ and $Y$ be the two transitive constituents of $G_t$. Further say that the letter $a_1$ of $G_t$ that $F$ fixes is found in the constituent $X$. Then $F$ fixes $2^\beta>1$ letters of $X$. Assume that $F$ fixes none of the letters of $Y$. In $G_t$ the largest group $N$ in which $F$ is invariant has a regular constituent on the $2^\beta$ letters fixed by $F$. Since $G_{t-1}$ is primitive and since $G_t$ has exactly two transitive constituents, $G_{t-1}$ contains a substitution* $S=(b_1a_1)\cdots$ which transforms $F$ into itself and hence permutes among themselves the remaining $2^\beta-1$ letters of $X$ that $F$ fixes. Consequently the group $\{N, S\}$ has a doubly transitive constituent of order $2^\beta(2^\beta+1)$ on the $2^\beta$ letters of $X$ and the letter $b_1$, for the subgroup that fixes $b_1$ of $\{N, S\}$ is in $G_t$ and hence is $N$.

The $G_{t-1}$ which fixes $b_2$ but displaces $b_1$ likewise contains a substitution $S'=(b_2a_1)\cdots$ which transforms $F$ into itself. The group $\{N, S, S'\}$ has a 3-ply transitive constituent of order $2^\beta(2^\beta+1)(2^\beta+2)$. Thus in the $t$-ply transitive group $G$ a group $N'=\{N, S, S', S'', \cdots\}$ can be found which has a $(t+1)$-ply transitive constituent of degree $2^\beta+t$ and of order $2^\beta(2^\beta+1)(2^\beta+2)\cdots(2^\beta+t)$. However, Jordan has shown that such a multiply transitive constituent is alternating or symmetric if $t>4$. Thus $2^\beta=2$, and the multiply transitive constituent of $N'$ is symmetric.

Now consider a substitution $(b_1b_2b_3)(b_4)(b_5)\cdots(b_t)(a_1)(a_2)\cdots$ of order three of $N'$, $a_1$ and $a_2$ being the two letters fixed by $F$. Since this substitution also transforms $G_t$ into itself, it must also fix at least one letter, $a_x$, of the constituent $Y$ of $G_t$. Thus it occurs in the $G_t$ which fixes $a_1, a_2, a_x, b_4, b_5, \cdots, b_t$.

* W. A. Manning, these Transactions, vol. 29 (1927), p. 815, §§1 and 5, Corollary II.

Hence $F$ fixes some letters from each transitive constituent of $G_t$. First, it will be shown that the transitive constituents of $G_t$ are simply isomorphic. Suppose them multiply isomorphic. Choose $F$ as the subgroup which fixes one letter of the constituent of smaller order. Then, contrary to the above statement, $F$ does not fix letters from each transitive constituent of $G_t$.

Further it will be shown that the two constituents are of the same degree. Let $F$ be the subgroup which fixes one letter of the constituent of smaller degree. The order of this constituent will be $wg_1$, where $w$ is the degree of the constituent and $g_1$ is the order of $F$. Since $F$ also fixes letters of the second constituent of $G_t$, the order $g_2$ of the subgroup that fixes one letter of this constituent is at least $g_1$. Let $m$ be the degree of the second constituent. Then $wg_1 = mg_2$, where $g_2 \geqq g_1$ and $m \geqq w$. Hence $g_1 = g_2$ and $w = m$.

Thus $F$ fixes $2^\beta$ letters from each transitive constituent of $G_t$ and $N$ has two simply isomorphic regular constituents on the letters which $F$ fixes. If none of the substitutions $S, S', \cdots$ connect the letters of the two regular constituents of $N$, the group $N'$ may be formed and the reasoning of the above paragraphs may be applied to its multiply transitive constituent. As before $2^\beta = 2$ and in this case the substitution $(b_1b_2b_3)(a \cdots) \cdots$ of order three of $N'$ fixes $b_4, b_5, \cdots, b_t$ and the 4 letters which $F$ fixes. Hence it occurs in a conjugate of $G_t$. Then at least one $S$ connects the sets of letters in question. Assume this substitution to be $S$. Then the group $\{N, S\}$ has a primitive constituent of degree $2^{\beta+1}+1$ and of class $2^{\beta+1}$ on the letters that $F$ fixes and the letter $b_1$. Hence this constituent is of degree $p^\alpha$, $p$ a prime, and contains a characteristic elementary subgroup.*

As before, form the group $N'$. Since the substitution $(b_1b_2b_3)(b_4)(b_5) \cdots (b_t)(a_1) \ldots$ of order three of $N'$ transforms both $G_t$ and $F$ into itself, it permutes among themselves the letters of each set of $2^\beta$ $a$'s that $F$ fixes. Hence it fixes exactly one $a$ from each of these sets, for if it fixed more than one letter from each set, this substitution would occur in a conjugate of $G_t$. Thus $2^\beta \equiv 1$, mod 3, and $p = 3$. The group $N'$ cannot have so much as a 5-ply transitive constituent in the letters in question, for let $T = (a_1a_2)(a_3a_4) \cdots$ be a substitution of order two of $N$ on these letters. Then in $N'$ there exist the substitutions

$$V = \begin{pmatrix} b_1a_1a_2a_3a_4b_6 \cdots \\ b_1b_2b_3b_4b_5b_6 \cdots \end{pmatrix} \text{ and } V^{-1}TV = (b_1)(b_2b_3)(b_4b_5)(a_xa_y) \cdots .$$

Note that the substitution $V^{-1}TV$ transforms the elementary group of degree $3^\alpha$ into itself. Now the partial substitution $U = (a_xa_y) \cdots$ fixes 5 letters of

* G. Frobenius, Berliner Sitzungsberichte, 1902, pp. 455–459.

the constituent of $N$ in question and transforms the elementary subgroup of $\{S, N\}$ of degree $3^\alpha$ into itself. This is obviously impossible. Hence $t \leqq 4$, contrary to the present hypothesis.

Then the subgroup $F$ reduces to the identity and $G_t$ is of class $n-t$. Since $G_{t-1}$ is primitive, it is of degree $p^\alpha$, $p$ a prime, and contains a characteristic elementary subgroup. Further since $G_t$ is of class $n-t$, each of its transitive constituents is of degree $2^\gamma$. Now $2^\gamma \equiv 1$, mod 3, for the substitution $(b_1b_2b_3)(b_4)\cdots(b_t)\cdots$ of order three of $I'$ fixes only one $a$ from each of the transitive constituents of $G_t$. Then $2^\gamma+2^\gamma+1\equiv 0$, mod 3, and $p=3$. Apply the reasoning of the previous paragraph to the groups $G$, $G_{t-1}$, and $G_t$, and it is seen that $t \leqq 4$.

Then since $t>5$, $I'$ permutes some, and hence $t$ or more, transitive constituents of $G_t$. Thus the following theorem has been proved:

THEOREM 2. *If $G_t$, the subgroup that fixes $t$ letters of a $t$-ply transitive group $G$, is of order $2^m$, the largest subgroup in which $G_t$ is invariant permutes $t(t>5)$ or more transitive constituents of $G_t$.**

5. At this point it is necessary to assume $t \geqq 8$, and this assumption will be held throughout the remaining part of this paper. It has been shown that the substitutions of $I'$ permute some transitive constituents of $G_t$. Hence $I'$ contains an imprimitive constituent, whose systems of imprimitivity are transitive constituents of $G_t$ and which is multiply isomorphic to the alternating group of degree $t$. Consider the group $H$ whose letters are the systems of imprimitivity of the imprimitive constituent just described. It is simply isomorphic to the alternating constituent $A$ of degree $t$ in $I'$. Let $A_1$ be the subgroup of $A$ which is simply isomorphic to $H_1$, the subgroup which fixes one letter of $H$. Now $A_1$ has either one transitive constituent of degree $>\frac{1}{2}t$, or all of its transitive constituents are of degree $\leqq \frac{1}{2}t$. Let the former be true, and call the transitive constituent of degree $\beta > \frac{1}{2}t$, $B$. Let $B$ contain the alternating group of degree $\beta$ of $A_1$ which fixes $t-\beta$ letters of $A$.

Under this hypothesis, it will be shown that $A_1$ cannot be transitive. Note that if it is transitive it is alternating. Let $A_1$ be transitive and of degree $t-1$. Then $H$ is of degree $t$ and is the alternating group. Thus a substitution of degree and order three of $A_1$ is associated† with a substitution of degree and order three of $H_1$. Since $t \geqq 8$, this substitution fixes more than three letters of $H$ and in $I'$ this substitution which has a cycle of three letters in $A_1$ permutes among themselves the letters of more than three transitive constituents of $G_t$. Now since the transitive constituents of $G_t$ are of degree

* Note that Theorem 2 holds also for $t=5$.

† M. J. Weiss, these Transactions, vol. 30 (1928). See footnote on p. 337.

a power of two, this substitution fixes more than three letters of $G_t$. Hence it fixes $t$ or more letters of $G$, and consequently $G_t$ contains a substitution of order three, contrary to the present hypothesis.

If $A_1$ is of degree $t-2$ and transitive, $H$ is of degree $t^2-t$ and its substitutions record the permutations of the $t^2-t$ sequences of two letters by the alternating group of degree $t$. Now a substitution of degree and order three of the alternating group of degree $t$ fixes exactly $(t-3)(t-4)$ sequences. Hence a substitution of $H_1$ which is associated with a substitution of degree and order three of $A_1$ fixes exactly $(t-3)(t-4)$ transitive constituents of $G_t$. Then a substitution of $I'$ which has a cycle of three letters in $A_1$ fixes at least $(t-3)(t-4)>2$ letters of $G_t$. Hence such a substitution fixes $t$ or more letters of $G$.

Now $A_1$ cannot be transitive of degree $t-3$ or less, for then it is invariant in a group of at least three times its order and hence $H_1$ fixes three or more transitive constituents of $G_t$. Then in this case $G_t$ also contains a substitution of order three. Thus $A_1$ is necessarily intransitive, containing an invariant alternating group of degree $\beta$, which fixes $t-\beta$ letters of $A$, and a constituent, transitive or intransitive, of degree $\leqq t-\beta$.

Moreover the constituent $B$ is of degree $<t-2$. If $B$ is of degree $t-2$, $A_1$ is of order $(t-2)!$. Then $H$ is of degree $(t^2-t)/2$ and records the permutations of the $(t^2-t)/2$ transpositions of the alternating group of degree $t$. A substitution of degree and order three of $A$ fixes exactly $(t-3)(t-4)/2$ transpositions. Hence the substitution of $H$ associated with it fixes $(t-3)(t-4)/2>2$ transitive constituents of $G_t$. Thus unless $B$ is of degree $<t-2$, $G_t$ contains a substitution of order three.

It is necessary to determine the order of $A_1$. First it will be shown that $A_1$ contains all the substitutions of degree and order three of $A$ which fix the letters of the constituent $B$. For this purpose it is necessary to have in mind Dyck's theorem* on transitive groups simply isomorphic to a given group. Now choose substitutions $S_j$ not in $A_1$ such that $A$ can be written in the array $A_1S_j$, $j=1, 2, \cdots, t!/(2m)$, where $m$ is the order of $A_1$. Denote the substitutions of $A_1$ by $\sigma_i$, $i=1, 2, \cdots, m$. Multiply this array on the right by the substitutions $\sigma_i$. Then by Dyck's theorem, to each substitution $\sigma_k$ of $A_1$, there corresponds the substitution

$$\begin{pmatrix} A_1S_j \\ A_1S_j\sigma_k \end{pmatrix} \qquad (j = 1,2, \cdots, t!/(2m)),$$

and the substitutions of $H$ may be regarded as written on the sets of letters

* W. Dyck, Mathematische Annalen, vol. 22 (1883), pp. 70–108.

$A_1S_j$. Thus the substitution of $H_1$ associated with the substitution $\sigma_k$ of $A_1$ is the above substitution.

Now suppose that $S$ is a substitution of degree and order three of $A$ which fixes the letters of the constituent $B$ and which is not found in $A_1$. The sets $A_1S$ and $A_1S^2$ found in the above array are distinct, for if these sets had a substitution in common, that is, if

$$\sigma_m S = \sigma_n S^2,$$

then

$$S = \sigma_n^{-1}\sigma_m,$$

but $S$ is not a substitution of $A_1$. Further let $\sigma$ be a substitution of degree and order three of $A_1$ on the letters of the constituent $B$ only. Then $\sigma$ and $S$ are commutative since they are written on different sets of letters. Hence to the substitution $\sigma$ corresponds the substitution

$$\begin{pmatrix} A_1S_j \\ A_1S_j\sigma \end{pmatrix} \qquad (j = 1,2, \cdots, t!/(2m)),$$

which fixes three of the sets $A_1S_j$, namely, $A_1S_1$, $A_1S$, and $A_1S^2$, where $S_1=1$. Thus since the letters of $H$ are also the transitive constituents of $G_t$, the substitution of order three of $I'$ which coincides with $\sigma$ in its cycle on the letters of $A_1$ fixes at least three transitive constituents of $G_t$, and hence is found in a $G_t$.

Thus $A_1$ also contains the alternating group on the $t-\beta$ letters not found in the constituent $B$. Hence the minimum order of $A_1$ is $\beta!(t-\beta)!/4$.

Now consider that subgroup $D$ of $A_1$ which is the direct product of the alternating group of degree $\beta$ and a substitution $T$ of degree and order three on the letters of the constituent of degree $t-\beta$. Since $A$ and $H$ are simply isomorphic groups, $H$ contains a subgroup $J$ which is simply isomorphic to the group $D$. Further the subgroup of $H_1$ which is simply isomorphic to the alternating group of degree $\beta$ displaces letters from every transitive constituent of $H_1$, for if it fixes the letters of one of the transitive constituents, a substitution of degree and order three of the alternating group of degree $\beta$ fixes at least three transitive constituents of $G_t$. The three transitive constituents fixed would be the one that $H_1$ fixes and the ones which are the letters of the transitive constituent of $H_1$ in question. Hence this substitution would fix at least three letters of $G_t$ and $G_t$ would contain a substitution of order three. Thus since the alternating group of degree $\beta(\beta \geqq 5)$ is simple, every transitive constituent of $H_1$ has a subgroup simply isomorphic to it and associated with it. Again for the above reason the substitution $T$ displaces letters of every transitive constituent of $H_1$. Hence every transitive constituent of $H_1$ has a subgroup simply isomorphic to $D$ and associated with $D$.

Let $J'$ be one of the transitive constituents of $J$ simply isomorphic to $D$. Consider the subgroup $D_1$ of $D$ which is simply isomorphic to the subgroup $J_1'$ which fixes one letter of $J'$. It will be shown that $D_1$ cannot contain a substitution of degree and order three. First it cannot contain $T$, for $T$ is invariant in $D$. Thus $D_1$ is invariant in a group of at least three times its order and consequently $J_1'$ fixes at least three letters of $J'$. Then the substitution of $J_1'$ associated with a substitution of degree and order three of $D_1$ fixes at least three transitive constituents of $G_t$, and $G_t$ contains a substitution of order three.

Since $D_1$ contains no substitution of degree and order three a theorem by Bochert* may be applied to determine its maximum order. Its order thus $\leqq 3\beta!/[(\beta+1)/2]!$, where $[q]$ denotes the integral part of $q$. Then a transitive constituent of $J$ is of degree $\frac{1}{3}[(\beta+1)/2]!$ at least.

The value of $\beta$ must now be determined so that the degree of $H$ can be evaluated. Under the present hypothesis, the order of $A_1 \leqq \beta!(t-\beta)!/2$, and hence the minimum degree of $H$ is $t!/\{\beta!(t-\beta)!\}$. The maximum degree of $H$ is found by determining the minimum order of $A_1$. The latter has been seen to be $\beta!(t-\beta)!/4$ and thus the maximum degree of $H$ is $2t!/\{\beta!(t-\beta)!\}$. When $t>8$, $H$ is simply transitive,† and $J$ must contain at least two transitive constituents. If $J$ and consequently $H_1$ contains only one transitive constituent, $H_1$ fixes exactly two letters of $H$, for it has been seen (§5, paragraph 3) that $H_1$ cannot fix so many as three letters. Then $H$ is imprimitive with systems of two letters. Its group in the systems is then either a doubly transitive group simply isomorphic to the alternating group of degree $t$ or the alternating group itself. The former case contradicts Maillet's theorems. In the latter case a substitution of degree and order three of $A_1$ has a substitution of degree and order three associated with it in the group in the systems of $H$. Consequently the substitution of $H_1$ associated with a substitution of degree and order three of $A_1$ fixes all except 6 letters of $H$. Then a substitution of $G_t$ which has a cycle of three letters in $A_1$ fixes more than three transitive constituents of $G_t$. This has been shown previously to be impossible.

Now the least degree of $J$ must always be less than the maximum degree of $H$. Hence

$$[(\beta+1)/2]! < 2t!/\{\beta!(t-\beta)!\}, \quad t > 8. \tag{1}$$

Note that when $t=8$, $\beta$ is determined, for $t/2<\beta\leqq t-3$. When $t>8$ and $\beta<t-3$, inequality (1) will be used to determine $\beta$.

Since the degree of each transitive constituent of $G_t$ is at least two, $G_t$

* A. Bochert, Mathematische Annalen, vol. 40 (1889), p. 584.

† E. Maillet, Journal de Mathématiques, (5), vol. 1 (1895), p. 5.

has at most $(n-t)/2$ transitive constituents. Now the maximum number of transitive constituents of $G_t$ must exceed or equal the minimum degree of $H$, namely,

$$(n-t)/2 \geqq t!/\{\beta!(t-\beta)!\}. \tag{2}$$

It remains to determine a lower bound for $(n-t)/2$. The right hand member of inequality (2) decreases as $\beta$ increases, $\beta > t/2$. Now it has been seen that $\beta$ must be chosen so that inequality (1) holds. Hence if $\beta(<t-3)$ is chosen to be one unit less than the least value of $\beta$ which satisfies the inequality

$$[(\beta+1)/2]! \geqq 2t!/\{\beta!(t-\beta)!\},\ t > 8, \tag{3}$$

inequality (2) holds for the value of $\beta$ thus determined.

The above discussion may be summarized as follows:

If $A_1$ has an alternating constituent of degree $>t/2$, inequality (2) gives the relation between $n$ and $t$, with $t/2 < \beta \leqq t-3$ or $\beta(<t-3)$ determined by inequality (3).

6. Now it may be assumed that the transitive constituent of degree $\beta(>t/2)$ of $A_1$ is not alternating. Let it be imprimitive. The largest imprimitive group of degree $\beta$ is of order $2(\beta/2)!^2$ or $2\{(\beta-1)/2\}!^2$, according as $\beta$ is even or odd. Hence the order of $A_1 \leqq [\beta/2]!^2\ (t-\beta)!$ and $H$ is of degree $t!/\{2[\beta/2]!^2(t-\beta)!\}$ at least. Thus in this case the following inequality holds:

$$(n-t)/2 \geqq t!/\{2[\beta/2]!^2(t-\beta)!\},\ t/2 < \beta \leqq t. \tag{4}$$

Let the constituent of degree $\beta$ be primitive. Since a non-alternating primitive group contains no substitution of degree and order three, the theorem of Bochert quoted above may be used to determine the order of the constituent of degree $\beta$. The order of $A_1 \leqq \beta!(t-\beta)!/[(\beta+1)/2]!$ and $H$ is of degree $t![(\beta+1)/2]!/\{2\beta!(t-\beta)!\}$ at least. Thus if the constituent of degree $\beta$ is primitive, the following inequality holds:

$$(n-t)/2 \geqq t![(\beta+1)/2]!/\{2\beta!(t-\beta)!\},\ t/2 < \beta \leqq t. \tag{5}$$

The minimum of each of the right hand members of inequalities (4) and (5) will now be determined. Denote them by $m_1(\beta)$ and $m_2(\beta)$, respectively. Recall Stirling's formula for $\log t!$, namely,

$$\log t! = (t+\tfrac{1}{2})\log t - t + \tfrac{1}{2}\log 2\pi + \theta/(12t),\ 0 < \theta < 1.$$

Then

$$\log m_1(\beta) = \left(t+\frac{1}{2}\right)\log t + \frac{\theta}{12t} - \log 2 - \left(2\left[\frac{\beta}{2}\right]+1\right)\log\left[\frac{\beta}{2}\right] + 2\left[\frac{\beta}{2}\right] - \log 2\pi - \frac{\theta_1}{6\left[\frac{\beta}{2}\right]} - \left(t-\beta+\frac{1}{2}\right)\log(t-\beta) - \beta - \frac{\theta_2}{12(t-\beta)}, \quad \frac{t}{2} < \beta \leqq t-1, \ 0<\theta_i<1.$$

The cases $\beta$ even and $\beta$ odd must be treated separately in order to apply Stirling's formula. However in both cases, the second derivative of $\log m_1(\beta)$ is negative. Hence the minimum of $\log m_1(\beta)$ occurs when $\beta$ assumes its least even or odd value or when $\beta$ assumes its greatest even or odd value, $[t/2]+1 \leqq \beta \leqq t-1$. Consequently the minimum of $m_1(\beta)$ occurs for some one of these values of $\beta$. It is found that if $t$ is odd, the minimum occurs when $\beta=t-1$. If $t$ is even, the minimum occurs when $\beta=t$. Thus the minimum of $m_1(\beta)$ is $t!/(2[t/2]!^2)$ and the inequality (4) becomes

$$(n-t)/2 \geqq t!/(2[t/2]!^2). \tag{6}$$

Apply Stirling's formula to $\log m_2(\beta)$. Then

$$\log m_2(\beta) = \left(t+\frac{1}{2}\right)\log t - \left(\beta+\frac{1}{2}\right)\log\beta - \left(\left[\frac{\beta+1}{2}\right]+\frac{1}{2}\right)\log\left[\frac{\beta+1}{2}\right] - \log 2 - \left(t-\beta+\frac{1}{2}\right)\log(t-\beta) - \left[\frac{\beta+1}{2}\right] + \frac{\theta}{12t} - \frac{\theta_1}{12\beta} + \frac{\theta_2}{12\left[\frac{\beta+1}{2}\right]} - \frac{\theta_3}{12(t-\beta)}, \quad \frac{t}{2} < \beta \leqq t-1, \ 0<\theta_i<1.$$

The cases $\beta$ even and $\beta$ odd must again be considered separately. However the second derivative of $\log m_2(\beta)$ is again negative in both cases. Hence the minimum of $\log m_2(\beta)$ occurs for the extreme values of $\beta$, and consequently the minimum of $m_2(\beta)$ occurs for some one of these extreme values of $\beta$. Consider the values of $\beta$ at the lower end of the interval. When $\beta$ is even the least value of $\beta$ is $(t+1)/2$, $(t+3)/2$, $(t+2)/2$, $(t+4)/2$, according as $t=2k+1$, $2k-1$, $2k$, or $4m$, $k$ odd. When $\beta$ is odd the least value of $\beta$ is $(t+3)/2$, $(t+1)/2$, $(t+4)/2$, $(t+2)/2$, according as $t$ is of one of the above forms. For all these values of $\beta$ except for $\beta=(t+3)/2$ and $t=9$, $m_2(\beta) \geqq m_1(t)$. However if $t=9$, the minimum of $m_2(\beta)$ occurs for $\beta=9$. Hence inequality (5) need not be considered for values of $\beta$ at the lower end of the $\beta$ interval.

Then only the values of $m_2(\beta)$ for $\beta=t-2$, $t-1$, and $t$ need be considered. Now $m_2(t)$ is less than $m_2(t-2)$ or $m_2(t-1)$. Thus inequality (5) may be replaced by the inequality

$$(n-t)/2 \geqq \tfrac{1}{2}[(t+1)/2]!. \tag{7}$$

7. If no transitive constituent of $A_1$ is of degree $>t/2$, the order of $A_1 \leqq \frac{1}{2}[t/2]!^2$ and for this case

$$(n-t)/2 \geqq t!/[t/2]!^2. \tag{8}$$

However since the right hand member of inequality (8) is greater than the right hand member of inequality (6), the former may be discarded.

8. Thus if $G_t$ is of order $2^m$ and $t \geqq 8$, the following theorem has been proved:

THEOREM 3. *Let the subgroup that fixes t letters of a t-ply transitive group of degree n and class $>3$ be of order $2^m$. Then if $t \geqq 8$,*

$$(n-t)/2 \geqq t!/\{\beta!(t-\beta)!\},$$

*where $\beta$ is an integer chosen such that $t/2 < \beta \leqq t-3$, or chosen to be one unit less than the least value of $\beta$ which satisfies the inequality*

$$[(\beta+1)/2]! \geqq 2t!/\{\beta!(t-\beta)!\};$$

*or*

$$n-t \geqq [(t+1)/2]!;$$

*or*

$$n-t \geqq t!/[t/2]!^2.$$

*The symbol $[s]$ denotes the integral part of s.*

When $t$ is sufficiently great, it will be shown that the last inequality in Theorem 3 is the only one necessary to consider. Comparing the logarithms of the right hand members of the second and third inequalities, it is found by means of Stirling's formula, that when $t \geqq 16$, the former is always greater than the latter. Hence the former may be discarded when $t \geqq 16$.

Further analyze the inequality (3) which limits $\beta$, and write it in the form

$$V(\beta) = [(\beta+1)/2]!\beta!(t-\beta)!/(2t!) \geqq 1.$$

Then if $\beta$ is so chosen that $\log V(\beta) \geqq 0$, the inequality (3) holds. Now by Stirling's formula, it is found that $\log V(\beta)$ is positive for $\beta=t/2$ as soon as $t \geqq 160$. Then for $t \geqq 160$, $\beta$ may be chosen equal to $t/2$, but $\beta > t/2$ by hypothesis. Hence when $t \geqq 160$ the first formula in Theorem 3 may also be discarded.

For later comparison purposes it is of interest to find the principal value of $\log (n-t)$ when $t$ becomes infinite. Now

$$\log (n - t) \geqq \log t! - 2 \log [t/2]!.$$

The right hand member of this inequality, when expanded by Stirling's formula, becomes

$$(t + \tfrac{1}{2}) \log t - (t + 1) \log (t/2) + \theta/(12t) - \theta_1/(3t) - \tfrac{1}{2} \log 2\pi = (t + 1) \log 2 - \tfrac{1}{2} \log t + \theta/(12t) - \theta_1/(3t) - \tfrac{1}{2} \log 2\pi,$$

where $0<\theta_i<1$, $i=1, 2$, the principal value of which is $t \log 2$. Thus

$$\log (n - t) \geqq t! \log 2(1 + \epsilon), \tag{9}$$

where $\epsilon$ approaches 0 as $t$ approaches $\infty$.

## $G_t$ OF ORDER $p^\alpha m$, $p$ AN ODD PRIME

9. In the remaining part of this paper it will be assumed that the order $g_t$ of the subgroup $G_t$ which fixes $t$ letters of the $t$-ply transitive group $G$ is divisible by an odd prime $p$. The analysis of this case is an extension of the method given by Jordan in his paper in Liouville's Journal of 1895.

Let $p$ be the greatest prime which divides the order of $G_t$. The following slight extension of a theorem by Jordan* will be needed in the development of the theory for the case under discussion:

Consider a Sylow subgroup $P$ of $G_t$ of order $p^\alpha$. Then $g_t=vp^\alpha(rp+1)$, where $vp^\alpha$ is the order of the largest subgroup of $G_t$ in which $P$ is invariant. Let $a_1, a_2, \cdots$ be the letters that $G_t$ displaces and $b_1, b_2, \cdots, b_t$ the $t$ letters that $G_t$ fixes. Since $G$ is $t$-ply transitive, there exists a substitution $S=(b_ib_k)(b_lb_m)(a\cdots)\cdots$ which transforms $G_t$ into itself, $b_i, b_k, b_l, b_m$ being any four of the above $t$ letters. In the group of order $2g_t$ thus obtained, let the order of the largest group in which $P$ is invariant be $v_1p^\alpha$. Then $v_1=2v$ or $v$ according as the new substitutions introduced do or do not transform $P$ into itself. Sylow's theorem shows that $v_1=2v$, for $v_1p^\alpha(r_1p+1)=2vp^\alpha(rp+1)$, from which it is found that $v_1\equiv 2v$, mod $p$. Then for each set of possible values of the subscripts $i, k, l, m$ there exists a substitution $S_j$ which transforms $P$ into itself. Now form the group $W=\{P, S_1, S_2, \cdots\}$. It contains $P$ invariantly and also $P$'s characteristic elementary subgroup $L$. Further $W$ has an alternating constituent on the $t$ letters that $G_t$ fixes. The above discussion may be summarized as follows:

* C. Jordan, Journal de Mathématiques, (5), vol. 1 (1895), p. 37.

*If the order of the subgroup which fixes t letters of a t-ply transitive group G is divisible by an odd prime p, a subgroup W of G can always be found which contains an invariant elementary subgroup L of order $p^\beta$ on the letters of $G_t$ and which has an alternating constituent on the t letters that $G_t$ fixes.*

The substitutions of $W$ are of the form $ABPC$, where $A$ denotes a substitution on the letters of the alternating constituent of degree $t$, $B$ a substitution which permutes the transitive constituents of $L$, $P$ the product of substitutions $P_1, P_2, \cdots$, each of which permutes among themselves the letters of a transitive constituent of $L$, and $C$ a substitution on the letters of $G_t$ not contained in $L$. The substitutions $P_1$, say, on the $p^\mu$ letters of a transitive constituent of $L$ either do or do not generate a subgroup of the holomorph of the elementary group of degree $p^\mu$ which has a quotient group simply isomorphic to the alternating group of degree $t$. If the former be true, Jordan has shown that

$$\mu \geqq t - 3 - \log t/\log 2.*$$

Now the degree of a transitive constituent of $L$ is $<n-t$. Hence

$$n - t > p^\mu,\ p > 2,\ \mu \geqq t - 3 - \log t/\log 2. \tag{10}$$

10. Suppose that this inequality does not hold. Then none of the sets of subsitutions $P_i$, $i=1, 2, \cdots$, generate a subgroup of the holomorph of the elementary group of degree $p^{\mu_i}$ which has a quotient group simply isomorphic to the alternating group of degree $t$. Hence $W$ contains a subgroup $\overline{W}$, with an alternating constituent of degree $t$, in which each of the sets of substitutions $P_i$, except those forming the transitive constituents of $L$ that are permuted in $W$, reduces to the identity. Now note that the elementary group $L$ is of degree $\geqq(n-1)/2$, $n$ being the degree of $G$, for the degree of $L$ must be equal to or greater than the class of $G$. Then substitutions of $\overline{W}$ must permute some transitive constituents of $L$, for, otherwise, a substitution which has a cycle of three letters in the alternating group of degree $t$ is of degree $\leqq 3+n-t-(n-1)/2<(n-1)/2$, the class of $G$.

Return to a consideration of the subgroup $W$. Since the substitutions of $\overline{W}$ permute some transitive constituents of $L$, the group $W$, which contains $\overline{W}$, also permutes some. The substitutions $B$ together with those of $L$ generate imprimitive groups whose systems of imprimitivity are the transitive constituents of $L$, and which are multiply isomorphic to the alternating group of degree $t$. Denote by $\mathfrak{B}_1, \mathfrak{B}_2, \cdots$, respectively, the groups in the systems of these imprimitive groups, and by $\mathfrak{C}_1, \mathfrak{C}_2, \cdots$ the groups generated by the substitutions $C$. Now the groups $\mathfrak{B}_i$ and $\mathfrak{C}_i$ are either all simply

* C. Jordan, Bulletin de la Société Mathématique de France, vol. 1 (1872), p. 55.

isomorphic to the alternating group of degree $t$ or some are not. Assume the latter to be true.

The groups $\mathfrak{B}_i$ and $\mathfrak{C}_i$ which are multiply isomorphic to the alternating group of degree $t$ then have invariant subgroups which fix the $t$ letters of the alternating constituent. Denote these invariant subgroups by $R$. Now all the substitutions $S_j=(b_ib_k)(b_lb_m)(a\cdots)\cdots$ of $W$ transform $R$ into itself. Then unless the order of $R$ is a power of two, choose its Sylow subgroup $P_1$ of order $p_1^{\alpha_1}$, $p_1$ an odd prime. The group $P_1$ reduces to $R$ if $R$ is of order a power of two. Then as in §9, paragraph 2, a group $W_1$ can be found in $W$ which transforms $P_1$ into itself and which has an alternating constituent on the $t$ letters that $G_t$ fixes. The substitutions $S_j$, $j=1, 2, \cdots$, in this case are written on the letters of the alternating constituent of degree $t$, on the transitive constituents of $L$ regarded as single symbols, and on the letters of $G_t$ itself.

Consider the characteristic elementary subgroup $L_1$ of $W_1$. The degree of none of the transitive constituents of $L_1$ exceeds $(n-t)/p$ or $(n-2t+1)/2$, the former formula giving the maximum degree of an imprimitive group whose systems of imprimitivity are the transitive constituents of $L$, the latter formula giving the maximum degree of a transitive constituent of $W$ which displaces none of the letters of $L$. Let $Z$ be the greater of these two quantities. The substitutions of $W_1$ are of the form $A\overline{B}B'PP'C$, where $\overline{B}$ denotes substitutions which permute transitive constituents of $L$, but which with the substitutions of $L$ generate imprimitive groups whose groups in the systems are simply isomorphic to the alternating group of degree $t$; $B'$ denotes substitutions which permute the transitive constituents of $L_1$; $P$ denotes the product of substitutions $P_1, P_2, \cdots$, each of which permutes among themselves the letters of a transitive constituent of $L$; $P'$ denotes the product of substitutions $P_1', P_2', \cdots$, each of which permutes among themselves the letters of a transitive constituent of $L_1$; and $C$ denotes substitutions on the letters of $G_t$ not found in $L$ or $L_1$. Recall that no one of the sets of substitutions $P_i$ generates a group which has a quotient group simply isomorphic to the alternating group of degree $t$.

If one of the sets of substitutions $P_i'$, $P_1'$, say, on the $p_1^{\gamma}$ letters of a transitive constituent of $L_1$, generates a subgroup of the holomorph of the elementary group of degree $p_1^{\gamma}$ which has a quotient group simply isomorphic to the alternating group of degree $t$, Jordan's inequality gives

$$Z \geqq p_1^{\gamma},\ p_1 \geqq 2,\ \gamma \geqq t - 3 - \log t/\log 2. \tag{11}$$

If this inequality does not hold, it may be assumed that none of the above sets of substitutions $P_i'$ generates a subgroup of the holomorph of an

elementary group which has a quotient group simply isomorphic to the alternating group of degree $t$. Then $W_1$ contains a subgroup $\overline{W}_1$ which has an alternating constituent of degree $t$ and in which each of the sets of substitutions $P_i'$, except those forming the transitive constituents of $L_1$ that are permuted by $W_1$, reduces to the identity. If $W_1$ fixes all the transitive constituents of $L_1$, the substitutions of $\overline{W}_1$ reduce to $A\overline{B}C$. This case will be considered later. Then assume that the substitutions of $W_1$ permute some transitive constituents of $L_1$.

Consider the imprimitive groups generated by the substitutions $B'$ and $L_1$. Denote their groups in the systems by $\mathcal{B}_1'$, $\mathcal{B}_2'$, $\cdots$, respectively, and the groups generated by the substitutions $C$, by $\mathcal{C}_1$, $\mathcal{C}_2$, $\cdots$, respectively. Again, unless all the groups $\mathcal{B}_i'$ and $\mathcal{C}_i$ are simply isomorphic to the alternating group of degree $t$, choose their invariant subgroups denoted by $R_1$ which fix the $t$ letters of the alternating group and repeat the above analysis, obtaining a group $W_2$ which contains an invariant elementary subgroup $L_2$. If all the groups $\mathcal{B}_i'$ and $\mathcal{C}_i$ are simply isomorphic to the alternating group of degree $t$, the analysis has been completed for the present purpose. The substitutions of $W_1$ then have the form $A\overline{B}\overline{B}'\overline{C}PP'$, where $A$, $\overline{B}$, $P$, $P'$ are defined as before; $\overline{B}'$ denotes substitutions which with those of $L_1$ generate imprimitive groups whose groups in the systems are simply isomorphic to the alternating group of degree $t$; and $\overline{C}$ denotes substitutions on the letters of $G_t$ not found in $L$ or $L_1$ which generate subgroups simply isomorphic to the alternating group of degree $t$. Note that none of the substitutions $P_i$ or $P_i'$ generate groups which have quotient groups simply isomorphic to the alternating group of degree $t$.

Now return to the case in which $W_1$ fixes all the transitive constituents of $L_1$, and consider the group $\overline{W}_1$, whose substitutions are of the form $A\overline{B}C$. If the substitutions $C$ generate groups which are simply isomorphic to the alternating group of degree $t$, the analysis has been completed. If they do not, take their invariant subgroups denoted by $\overline{R}_1$, which fix the $t$ letters of the alternating group, and apply the reasoning of the previous paragraphs, again obtaining a group $\overline{W}_2$ which contains an invariant elementary subgroup $L_2$.

Continue with this analysis until either a group $W_j$ or $\overline{W}_j$ of the desired type is obtained, or an inequality, such as inequalities (10) or (11), derived from this analysis holds. Obviously these two inequalities are the most unfavorable of any which might be obtained in carrying this analysis further. Consequently any further inequalities may be neglected. Then it remains to investigate the case when a group $W_j$ or $\overline{W}_j$ is obtained in which the substitutions are of the form

$$A\overline{B}\overline{B}'\overline{B}'' \cdots \overline{C}PP'P'' \cdots,$$

where the substitutions $\overline{B}$, $\overline{B}'$, $\overline{B}''$, $\cdots$ denote substitutions which permute transitive constituents of $L$, $L_1$, $L_2$, $\cdots$, respectively, but which with the substitutions of the latter generate imprimitive groups whose groups in the systems are simply isomorphic to the alternating group of degree $t$; $\overline{C}$ denotes substitutions on the letters of $G_t$ not found in $L$, $L_1$, $L_2$, $\cdots$, which generate groups simply isomorphic to the alternating group of degree $t$; $P, P', P'', \cdots$ denote the product of substitutions $P_1, P_2, \cdots, P_1', P_2', \cdots, P_1'', P_2'', \cdots$, each of which permute the letters of a transitive constituent of $L$, $L_1$, $L_2$, $\cdots$, respectively, among themselves, but which do not generate groups which have quotient groups simply isomorphic to the alternating group of degree $t$.

11. Thus suppose that $W_j$ is the first group in the series $W_i$, $\overline{W}_i$, $i=1, 2, \cdots$, of the desired type. Let the groups in the systems of the imprimitive groups generated by $\overline{B}^m$ and $L_m$ be denoted by $\mathfrak{B}_v^m$, $v=1, 2, \cdots$, and $m=0, 1, \cdots$, $\mathfrak{B}_v^0=\mathfrak{B}_v$, and the groups generated by the substitutions $\overline{C}$, by $\mathfrak{C}_v$. Recall that the groups $\mathfrak{B}_v^m$ and $\mathfrak{C}_v$ are all simply isomorphic to the alternating group of degree $t$. Consider a subgroup $K$ of $W_j$ which has an alternating constituent on the $k$ letters $b_1, b_2, \cdots, b_k$, where $5 \leqq k \leqq t$. First suppose that $K$ fixes more than one half the letters in each of the groups $\mathfrak{B}_v^m$ and $\mathfrak{C}_v$. Now all these letters or symbols fixed cannot reduce to the identity in $G_t$, for then the degree of a substitution of $W_j$ which has a cycle of three letters in the alternating constituent of degree $k$ is of degree $\leqq 3+(n-t)/2<(n-1)/2$. Then according to the theorem by Manning quoted in §3, $G$ is alternating. Therefore corresponding to at least one symbol fixed and arising from the letters of $L$, $L_1$, or $L_2$, $\cdots$, there exists a group simply or multiply isomorphic to the alternating constituent of degree $k$ which transforms an elementary group into itself. Now the degree of this group cannot exceed the degree of one of the systems of imprimitivity of an imprimitive group generated by the substitutions $\overline{B}^m$ and those of $L_m$. It has been shown that such a group permutes its systems according to a group simply isomorphic to the alternating group of degree $t$. Hence it has at least $t$ systems and the degree of a system is at most $(n-t)/t$. If this imprimitive group is a group generated by the substitutions $\overline{B}$ and those of $L$, Jordan's inequality gives

$$(12) \qquad (n-t)/t \geqq p^\delta,\ p>2,\ 5 \leqq k \leqq t-1,\ \delta \geqq k-3-\log k/\log 2.$$

If the substitutions $\overline{B}$ reduce to the identity in $K$, let the group under discussion be a group generated by the substitutions $\overline{B}'$ and those of $L_1$. The degree of this imprimitive group is then at most $Z$. Hence the degree of one of its systems is at most $Z/t$. In this case Jordan's inequality gives

(13) $Z/t \geqq p_1^{\delta_1}$, $p_1 \geqq 2$, $5 \leqq k \leqq t-1$, $\delta_1 \geqq k-3-\log k/\log 2$.

If the substitutions $\overline{B}'$ reduce to the identity in $K$, let the group in question be a group generated by the substitutions $\overline{B}''$ and those of $L_2$. Any further inequalities thus obtained evidently give more favorable bounds for the transitivity of $G$ and hence may be discarded.

12. It may now be assumed that the group $K$ displaces a half or more than a half of the letters in at least one of the groups simply isomorphic to the alternating group of degree $t$. Let $Q$ be the particular group in which $K$ displaces a half or more than one half of the letters. Let $\Gamma_1$ be the subgroup of the alternating constituent $\Gamma$ of degree $t$ which corresponds to the subgroup $Q_1$ that fixes one letter of $Q$. Now the degree of $Q$ is at most $(n-t)/p$ or $(n-2t+1)/2$, according as $Q$ has arisen from the letters of $L$ or from the letters of $G_t$ not found in $L$. On the other hand, the degree of $Q$ is $t!/(2q)$ where $q$ is the order of the subgroup $\Gamma_1$. Hence for this case the inequality

$$Z \geqq t!/(2q)$$

must be studied.

Let $\gamma_1, \gamma_2, \cdots$ be the degrees of the transitive constituents of $\Gamma_1$. Then the order of each transitive constituent group of $\Gamma_1$ divides $\gamma_i!$, respectively. Hence the order $q$ divides $\gamma_1!\gamma_2!\cdots$.

Recall the present hypothesis that the group $K$ fixes at most one half the letters in the group $Q$. Let $Q_1$ have $v$ conjugates under $Q$. Then $\Gamma_1$ has $v$ conjugates under $\Gamma$. Now those conjugates of $Q_1$ which fix letters that $K$ fixes, but which fix no letters that $K$ displaces, are invariant under $K$. Hence at most one half of the conjugates of $Q_1$ are invariant under $K$ and consequently at most one half the conjugates of $\Gamma_1$ are invariant under the alternating group of degree $k$ on the letters $b_1, b_2, \cdots, b_k$. Using this hypothesis Jordan* has shown that if one of the numbers $\gamma_1$ is greater than the number $\delta$, defined to be the greatest integer less than the quantity $t-(t-k+1)\cdot \log 2/(k+\log 2)$, the corresponding transitive constituent group of $\Gamma_1$ cannot contain the alternating group of the same degree.

Thus there are the following three cases to consider:

I. All the numbers $\gamma_i \leqq \delta$.

II. One number $\gamma_1 > \delta$, and the corresponding transitive constituent of degree $\gamma_1$ of $\Gamma_1$ is imprimitive.

III. One number $\gamma_1 > \delta$, and the corresponding transitive constituent of degree $\gamma_1$ of $\Gamma_1$ is primitive.

In Case I, the order $q \leqq \frac{1}{2}\delta!(t-\delta)!$. In Case II recall that the largest imprimitive group of degree $\gamma_1$ is of order $2(\gamma_1/2)!^2$ or $2((\gamma_1-1)/2)!^2$ ac-

* C. Jordan, Journal de Mathématiques, (5), vol. 1 (1895), p. 51.

cording as $\gamma_1$ is even or odd. Thus $q \leqq [\gamma_1/2]!^2(t-\gamma_1)!$ where $[s]$ denotes the integral part of $s$. In Case III, Bochert's theorem may be used to determine the order of the primitive constituent, and hence $q \leqq (t-\gamma_1)!\gamma_1!/[(\gamma_1+1)/2]!$. These three cases thus lead to the following inequalities:

$$Z \geqq t!/\{\delta!(t-\delta)!\}, \tag{14}$$

$$Z \geqq t!/\{2[\gamma_1/2]!^2(t-\gamma_1)!\}, \tag{15}$$

$$Z \geqq t![(\gamma_1+1)/2]!/\{2\gamma_1!(t-\gamma_1)!\}. \tag{16}$$

It has been seen (§6) that the minimum of the right hand member of the inequality (15) occurs when $\gamma_1=t$, and that inequality (16) need be considered only for $\gamma_1=t$. Further $Z$ may be taken equal to $(n-2t+1)/2$, for the greatest value of $(n-t)/p$ occurs when $p=3$, and the former number is always greater than the latter as soon as $n \geqq 4t-3$. Now $n \geqq 4t-3$, for the right hand members of the above inequalities are always $\geqq t$. Thus inequalities (14), (15), and (16) may be replaced by

$$(n-2t+1)/2 \geqq t!/\{\delta!(t-\delta)!\}, \tag{17}$$

$$(n-2t+1)/2 \geqq t!/\{2[t/2]!^2\}, \tag{18}$$

$$(n-2t+1)/2 \geqq \tfrac{1}{2}[(t+1)/2]!. \tag{19}$$

In the list of inequalities in §§9, 10, and 11 substitute the least possible values for $p$, $p_1$, and $Z$. Then only the following inequalities need be considered:

$$n-t \geqq 3^\alpha,\ \alpha \geqq t-3-\log t/\log 2, \tag{20}$$

$$(n-2t+1)/2 \geqq 2^\alpha, \tag{21}$$

$$(n-t)/t \geqq 3^\beta,\ \beta \geqq k-3-\log k/\log 2,\ 5 \leqq k \leqq t-1, \tag{22}$$

$$(n-2t+1)/(2t) \geqq 2^\beta. \tag{23}$$

In inequality (20), $\alpha>3$, for the holomorph of an elementary group of degree 9 or 27 obviously has no quotient group simply isomorphic to the alternating group of degree $t(\geqq 8)$. Likewise in inequality (21), $\alpha>3$. For a similar reason $\beta>3$ in inequalities (22) and (23), for in these cases the alternating group of degree $t$ is replaced by one of degree $k$. Although these same restrictions on $\alpha$ and $\beta$ might not hold if larger primes had been substituted for $p$ and $p_1$, similar considerations show that greater values for $p$ and $p_1$ give still more favorable values for $n$.

Now write inequalities (22) and (23) in the forms $n \geqq 3^\beta t+t$ and $n \geqq 2^{\beta+1}t+2t-1$, respectively. Since $3^\beta \geqq 2^{\beta+1}+1-1/t$ for all values of $\beta>3$, inequality (22) may be discarded. Again write inequalities (20) and (21) in

the forms $n \geqq 3^\alpha + t$ and $n \geqq 2^{\alpha+1} + 2t - 1$, respectively. Then if $2^{\alpha+1} + t - 1 < 3^\alpha$, inequality (20) may be discarded. Now $t - 1 < 2^\alpha$, $\alpha > 3$. Since $2^\alpha < 3^{\alpha-1}$ for all values of $\alpha > 3$, the above inequality holds.

Further, compare the right hand members of inequalities (18) and (19) with the right hand member of inequality (21). The logarithm of the right hand member of inequality (18) is always greater than the logarithm of the right hand member of inequality (21). Hence inequality (18) may be discarded. Again the logarithm of the right hand member of inequality (19) is greater than the logarithm of the right hand member of inequality (21) except when $t = 8$, $\alpha = 4$. However, when $t = 8$, inequality (17) gives a still lower bound for $n$ than inequality (19). Thus inequality (19) may be discarded for all values of $t$.

13. The results of the study of the $t$-ply transitive group $G$, the order of whose subgroup that fixes $t$ letters is divisible by an odd prime, may be summarized in the following theorem:

THEOREM 4. *Let the order of the subgroup that fixes $t$ letters of a $t$-ply transitive group of degree $n$ and class $> 3$ be divisible by an odd prime. Then if $t \geqq 8$,*

$$(n - 2t + 1)/2 \geqq 2^\alpha, \ \alpha \text{ an integer} > 3 \text{ and } \geqq t - 3 - \log t/\log 2;$$

*or*

$$(n - 2t + 1)/2 \geqq 2^\gamma t, \ \gamma \text{ an integer} > 3 \text{ and } \geqq k - 3 - \log k/\log 2,$$

*where $k$ is an integer such that $5 \leqq k \leqq t - 1$; or*

$$(n - 2t + 1)/2 \geqq t!/\{\delta!(t - \delta)!\},$$

*where $\delta$ is the greatest integer less than the quantity $t - (t - k + 1)\log 2/(k + \log 2)$.*

It may be of interest to find the principal value, when $t$ becomes infinite, of $(n - 2t + 1)/2$, or, which is more convenient, of $\log((n - 2t + 1)/2)$. The logarithm of the right hand member of the first inequality in Theorem 4 obviously has the principal value $t \log 2$. It remains to find the principal values of the right hand members of the second and third inequalities in the above theorem. Denote the logarithm of the right hand member of the second inequality by $f(k)$ and that of the third by $g(k)$. Thus

$$f(k) = (k - 3)\log 2 - \log k + \log t,$$

and by Stirling's formula

$$g(k) = (t + \tfrac{1}{2})\log t - (\delta + \tfrac{1}{2})\log \delta - (t - \delta + \tfrac{1}{2})\log(t - \delta) - \tfrac{1}{2}\log 2\pi + \theta/(12t) - \theta_1/(12\delta) - \theta_2/(12(t - \delta)), \text{ where } 0 < \theta_i < 1, \ i = 0,1,2.$$

Now $f(k)$ is a function which increases in value with $k$, while $g(k)$ is a function which decreases in value as $k$ increases. First determine $k$ so that the principal values of these functions are equal. Assume $k$ and $t$ of equal order. Then the principal value of $t-\delta$ is $\{(t-k)\log 2\}/k$. Now in $g(k)$ replace $\log \delta$ by

$$\log t + \log\{1 - (t-\delta)/t\} = \log t - (t-\delta)/t - (t-\delta)^2/(2t^2) - \cdots .$$

Then the principal value of $g(k)$ is $\{(t-k)\log 2 \log t\}/k$, while the principal value of $f(k)$ is obviously $k \log 2$. Setting these two values equal, it is found that

$$k = -\tfrac{1}{2}\log t + \tfrac{1}{2}(4t \log t + \log^2 t)^{1/2},$$

of which the principal value is $(t \log t)^{1/2}$. Thus $k$ is of order less than the order of $t$, and using this value of $k$, the principal value of $\log((n-2t+1)/2)$, derived from the second and third inequalities in the above theorem, is $(t \log t)^{1/2} \log 2$. This result is evidently less favorable than the one derived from the first inequality, and hence for the case $G_t$ of order $p^s m$, $p$ an odd prime, the following relation holds:

$$\log((n-2t+1)/2) \geqq (t \log t)^{1/2} \log 2(1-\eta), \tag{24}$$

where $\eta$ approaches 0 as $t$ approaches $\infty$.

In §8, it was shown that for the case $G_t$ of order $2^r$ the principal value of $\log(n-t)$ is $t \log 2$. Evidently this result gives a more favorable relation between $n$ and $t$ than the one just established. Hence inequality (24) holds for all $t$-ply transitive groups of class $>3$. This inequality is similar to the one found by Jordan in his 1895 paper, namely,

$$\log(n-t) \geqq (t \log t)^{1/2} \log 2(1-\epsilon),$$

where $\epsilon$ approaches 0 as $t$ approaches $\infty$.

UNIVERSITY OF CHICAGO,
CHICAGO, Ill.

# PROPERTIES OF SOLUTIONS OF AN INFINITE SYSTEM OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS OF THE FIRST ORDER WITH AUXILIARY BOUNDARY CONDITIONS*

BY

WILLIAM T. REID

## 1. Introduction

It is the purpose of this paper to consider an infinite system of ordinary linear differential equations of the first order. We write the system considered in the form

$$\frac{dy_i}{dx} = \sum_{\alpha=1}^{\infty} A_{i\alpha}(x) y_\alpha \qquad (i = 1,2, \cdots), \tag{1}$$

where $y(x) \equiv (y_i(x))$ is a vector in Hilbert space, i.e., $\sum_{\alpha=1}^{\infty} y_\alpha^2$ converges and $A(x) \equiv (A_{ij}(x))$ is a matrix each of whose elements is a Lebesgue summable function on $X: 0 \leqq x \leqq 1$. Furthermore, the matrix $A(x)$ is limited, in the sense defined by Hilbert, by a Lebesgue summable function $\phi(x)$ on $X$. In §2 preliminary definitions are given and the system of notation used throughout the paper is explained.

In §3 the properties of matrices of functions which satisfy system (1) and the corresponding adjoint system are considered. The adjoint system is given by

$$\frac{dz_i}{dx} = -\sum_{\alpha=1}^{\infty} A_{\alpha i}(x) z_\alpha \qquad (i = 1,2, \cdots). \tag{2}$$

In §4 we treat a system in which the matrix $A$ of the system (1) is a function of a parameter $\mu$, and sufficient conditions for the solution $y(x; \mu)$ to be continuous in $(x, \mu)$ are given. The differentiability of the solution with respect to the parameter $\mu$ is also considered.

In §5 we set up the system consisting of the equations of (1), together with boundary conditions that involve the value of the solution at two points. The set of boundary conditions that are adjoint to these is determined in a manner similar to that used by Bounitzky† for a finite system of ordinary linear differential equations of the first order.

* Presented to the Society, December 1, 1928, and March 29, 1929; received by the editors in June, 1929.

† E. Bounitzky, *Sur la fonction de Green des équations différentielles linéaires ordinaires*, Journal de Mathématiques, (6), vol. 5 (1909), pp. 65–125.

There is treated in §6 an infinite system in which each element of the matrix $A$ is a function of a parameter $\mu$ and $A(x;\mu)$ is limited by a summable function $\phi(x)$ uniformly with respect to $\mu$; furthermore, for each value of $\mu$ and for values of $x$ "almost everywhere" on $X$ the matrix $A(x;\mu)$ is completely continuous or "vollstetig," as defined by Hilbert.* This system, together with a special class of two-point boundary conditions, is shown to have many properties of a finite system.

Whenever the differential system of §6 is incompatible, it is possible to define a Green's matrix for the system. In §7 the existence and uniqueness of the Green's matrix is proved. We also state a theorem on defining a Green's matrix for values of the parameter which are characteristic numbers of the system. This theorem is more general than a theorem which has recently been established by W. M. Whyburn† for a finite differential system.

Boundary value problems for a finite system of ordinary linear differential equations of the first order whose coefficients are continuous, have been considered by Birkhoff and Langer‡ and by Bliss.§ The results of §§5 and 6 are an extension to an infinite system of some of the results obtained by Birkhoff and Langer and by Bliss for finite systems. More recently W. M. Whyburn‖ has considered a finite system of linear equations whose coefficients are Lebesgue summable functions and has shown that the solution of the differential system is the limit of the solution of the associated algebraic system.

W. L. Hart¶ has treated infinite systems of differential equations in four papers. The system considered in the first paper is of a different nature from the system here considered since the variables are defined in a more general

---

* D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, Berlin, 1912, p. 147.

† W. M. Whyburn, *On the Green's function for systems of differential equations*, Annals of Mathematics, (2), vol. 28 (1927), pp. 291–300.

‡ G. D. Birkhoff and R. E. Langer, *The boundary problems and developments associated with a system of ordinary linear differential equations of the first order*, Proceedings of the American Academy of Arts and Sciences, vol. 58 (1923), pp. 51–128.

§ G. A. Bliss, *A boundary value problem for a system of ordinary linear differential equations of the first order*, these Transactions, vol. 28 (1926), pp. 561–584.

‖ W. M. Whyburn, *On related difference and differential systems*, American Journal of Mathematics, vol. 51 (1929), pp. 265–286.

¶ W. L. Hart, *Differential equations and implicit functions in infinitely many variables*, these Transactions, vol. 18 (1917), pp. 125–160; *Linear differential equations in infinitely many variables*, American Journal of Mathematics, vol. 39 (1917), pp. 407–424; *The Cauchy-Lipschitz method for infinite systems of differential equations*, American Journal of Mathematics, vol. 43 (1921), pp. 226–231; *Functions of infinitely many variables in Hilbert space*, these Transactions, vol. 23 (1922), pp. 30–50.

space than Hilbert space. In the third paper it is proved by an extension to infinite systems of the notion of Cauchy polygons, and without the assumption of a Lipschitz condition, that the system of equations considered in the first paper has at least one solution satisfying arbitrary initial conditions. In the second paper an infinite linear system, in which the matrix of coefficients is a limited matrix and is analytic in the independent variable, is treated. The results there obtained are special cases of some of the results we obtain in §§3 and 4. In the fourth paper existence theorems for infinite differential equations in Hilbert space are given. However, system (1) that we consider does not come under Hart's theorem since we suppose that the elements of the matrix of coefficients are only Lebesgue summable functions. On page 408 of Hart's second paper references to earlier literature on infinite differential systems are given. More recently Feenberg* has treated an infinite system of equations in abstract space. Ettlinger† has considered an infinite system of the form (1) in which the elements of the matrix $A(x)$ are summable and in which the matrix $A(x)$ satisfies a condition which we show to be equivalent to the condition that is imposed in this paper. The results of this paper extend those that have been obtained by Ettlinger.

I desire to express my indebtedness to Professor H. J. Ettlinger who suggested to me the topic of this paper and who has assisted in its preparation by many suggestions and criticisms.

## 2. Preliminary definitions and notation

An infinite set of real numbers $\xi_1, \xi_2, \cdots$ represents a point or a vector in Hilbert real space provided the sum $\sum_{\alpha=1}^{\infty}\xi_\alpha^2$ converges. If $\xi \equiv (\xi_\alpha)$ and $\eta \equiv (\eta_\alpha)$ are two vectors the dot product of $\xi$ and $\eta$ is the sum $\sum_{\alpha=1}^{\infty}\xi_\alpha\eta_\alpha$ and is denoted by $\xi\eta$. The quantity $(\sum_{\alpha=1}^{\infty}\xi_\alpha^2)^{1/2}$ is the length or modulus of the vector $\xi$ and is denoted by $|\xi|$.

If $M$ is an infinite matrix we denote by $M_{ij}$ the element in the $i$th row and the $j$th column of $M$. All matrices considered in this paper, unless otherwise specified, are supposed to be real and to have an infinite number of rows and columns. When we say that a matrix or a vector has a property, which is not essentially a matrix or a vector property, we mean that each element of the matrix or each component of the vector has that property.

We consider in particular matrices that are limited matrices, as defined by Hilbert.‡

* For abstract of paper, see Bulletin of the American Mathematical Society, vol. 34 (1928), p. 709.

† For abstract of paper, see Bulletin of the American Mathematical Society, vol. 34 (1928), p. 708.

‡ D. Hilbert, loc. cit., p. 125.

**Definition.** A matrix $M \equiv (M_{ij})$ is said to be limited if there exists a constant $K$ such that for every pair of points $\xi$ and $\eta$ of Hilbert space and for every integer $n$,

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} \xi_i \eta_j \right| \leqq K \left( \sum_{i=1}^{n} \xi_i^2 \right)^{1/2} \left( \sum_{j=1}^{n} \eta_j^2 \right)^{1/2}.$$

When such a constant exists we say that the matrix $M$ is limited by the constant $K$. Properties of limited matrices have been considered by Hellinger and Toeplitz.* We will use the notation "H.T., page ——" to refer to a page in this article.

If $M \equiv (M_{ij})$ and $N \equiv (N_{ij})$ are two limited matrices, the sum matrix $M+N$ is given by $M+N \equiv (M_{ij}+N_{ij})$ and is also a limited matrix. If $k$ is a scalar and $M$ is a matrix, then $kM \equiv (kM_{ij})$. In particular, the zero matrix is the matrix each of whose elements is zero, and $E$ is the unit matrix $E \equiv (E_{ij})$, where $E_{ij}=0$ if $i \neq j$ and $E_{ii}=1$. If $M$ is a matrix the adjoint matrix $\overline{M}$ is defined by $\overline{M} \equiv (\overline{M}_{ij})$, where $\overline{M}_{ij} = M_{ji}$. We denote the vector $v \equiv (v_i)$, where $v_i = M_{i\alpha}\xi_\alpha$, by $M\xi$. In this expression $\alpha$ is an umbral subscript. That is, $M_{i\alpha}\xi_\alpha = M_{i1}\xi_1 + M_{i2}\xi_2 + \cdots$. Whenever it is possible to do so without causing confusion, umbral subscripts will be used. Similarly, if $\xi$ is any vector in Hilbert space the vector $u \equiv (u_i)$, where $u_i = M_{\alpha i}\xi_\alpha$, will be denoted by $\xi M$. If $M$ and $N$ are limited matrices, then the product $MN$ is given by $MN \equiv (M_{i\alpha}N_{\alpha j})$. Since it is shown by Hellinger and Toeplitz that each row and each column of a limited matrix is a vector in Hilbert space, the matrix $MN$ is well-defined and is also a limited matrix. In general $MN \neq NM$.

Let $v^{(1)}, v^{(2)}, \cdots$ be an infinite sequence of vectors, where $v^{(n)} \equiv (v_i^{(n)})$. If there exists a vector $v \equiv (v_i)$ such that $\lim_{n\to\infty} v_i^{(n)} = v_i$ $(i=1, 2, \cdots)$ we will say that the sequence of vectors converges to the vector $v$ and write $\lim_{n\to\infty} v^{(n)} = v$. If there exists a vector $v$ so that $\lim_{n\to\infty} |v^{(n)} - v| = 0$ we will say that the sequence of vectors *converges strongly*† to the limit vector $v$ and write $\mathrm{LIM}_{n\to\infty} v^{(n)} = v$.

Let $v(x)$ for each value of the variable $x$ on $a \leqq x \leqq b$ denote a vector in Hilbert space. If $x_0$ is a point on the interval $(a, b)$ and for every sequence of points $x_1, x_2, \cdots$ on this interval having $x_0$ as a sequential limit point we have $\mathrm{LIM}_{n\to\infty} v(x_n) = v(x_0)$, we say that the vector $v(x)$ is strongly continuous at $x = x_0$. If $v(x)$ is strongly continuous at every point on $(a, b)$, then $v(x)$ is said to be strongly continuous on this interval.

* E. Hellinger and O. Toeplitz, *Grundlagen für eine Theorie der unendlichen Matrizen*, Mathematische Annalen, vol. 69 (1910), pp. 289–330.

† The idea of strong convergence was introduced by Hilbert. See Hilbert, loc. cit., p. 127. See also F. Riesz, *Les Systèmes d'Équations Linéaires à une Infinité d'Inconnues*, Paris, 1913, p. 59.

Vitali* has proved the following theorem which will be used frequently in this paper.

THEOREM 2.1. *Let $f_1(x)$, $f_2(x)$, $\cdots$ be a sequence of Lebesgue summable functions of the bounded point set $X$ and such that $\lim_{n\to\infty} f_n(x) = f(x)$ on $X_0$.† If the absolute continuity of the integrals $\int f_n(x)dx$ is uniform on the set $X$, then $f(x)$ is a Lebesgue summable function on $X$ and*

$$\lim_{n\to\infty} \int_X f_n(x)dx = \int_X f(x)dx.$$

We now state without proof the following corollaries to this theorem:

COROLLARY 1. *Under the above hypotheses $\int_T f_n(x)dx$ converges to $\int_T f(x)dx$ uniformly for all measurable subsets $T$ of $X$. That is, for every $\epsilon > 0$ there is a positive $N(\epsilon)$ such that if $n > N(\epsilon)$ then for every measurable subset $T$ of $X$ we have*

$$\left| \int_T f_n(x)dx - \int_T f(x)dx \right| \leqq \epsilon.$$

COROLLARY 2. *If $\sum_{\alpha=1}^{\infty} A_\alpha(x)$ is a series of absolutely continuous functions which converges on the interval $X: a \leqq x \leqq b$ and the series*

$$\sum_{\alpha=1}^{\infty} \frac{d}{dx} A_\alpha(x)$$

*converges on $X_0$, and furthermore the absolute continuity of the integrals*

$$\int \sum_{\alpha=1}^{n} \frac{d}{dx} A_\alpha(x)dx$$

*is uniform with respect to $n$, then $\sum_{\alpha=1}^{\infty} A_\alpha(x)$ is an absolutely continuous function on $X$ and*

$$\frac{d}{dx}\left[ \sum_{\alpha=1}^{\infty} A_\alpha(x) \right] = \sum_{\alpha=1}^{\infty} \frac{d}{dx} A_\alpha(x) \text{ on } X_0.$$

Ettlinger‡ has recently stated a theorem on the continuity of a function

---

* M. Vitali, *Sull'integrazione per serie*, Rendiconti del Circolo Matematico di Palermo, vol. 23 (1907), pp. 137–155.

† $X_0$ is used to denote "almost everywhere" on $X$, or $X$ minus a null set. It is used in the generic sense and the excepted null set may vary.

‡ H. J. Ettlinger, *On continuity in several variables*, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 37–38. See also *Existence theorems for implicit functions of real variables*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 315–316.

of several variables. A theorem that is an extension of that of Ettlinger, and follows as a result of Corollary 1 above, will now be stated.

THEOREM 2.2. *Hypotheses*: (a) $f(x, y)$ *is summable on* $X: a \leqq x \leqq b$ *for each value of* $y$ *on* $Y: Y_1 < y < Y_2$; (b) *the absolute continuity of* $\int f(x, y)dx$ *is uniform on* $X$ *for* $y$ *on* $Y$; (c) *for values of* $x$ *on* $X_0$, $f(x, y)$ *is a continuous function of* $y$ *on* $Y$.

*Conclusion*: $F(x, y) = \int_a^x f(t, y)dt$ *is continuous in* $y$, *uniformly with respect to* $x$.

## 3. MATRIX SOLUTIONS OF AN INFINITE DIFFERENTIAL SYSTEM

According to the vector notation introduced in §2 we may write system (1) as

$$y'(x) = A(x)y(x). \tag{3.1}$$

Corresponding to this homogeneous system we have the non-homogeneous system

$$y'(x) = A(x)y(x) + b(x), \tag{3.2}$$

where $b(x)$ is an arbitrary vector which is, together with its length, summable on the interval of definition $X: 0 \leqq x \leqq 1$. The differential system we may also write in the alternative integral form

$$y(x) = \int_0^x [A(t)y(t) + b(t)]dt + y(0). \tag{3.2'}$$

A solution of (3.2) is defined as an absolutely continuous vector whose length is bounded uniformly on $X$ by a finite constant and which satisfies the vector differential equation (3.2) on $X_0$.

With the condition on the matrix $A(x)$ that there exist a non-negative summable function $\phi(x)$ on $X$ such that for every vector $\xi$ of Hilbert space $|A(x)\xi| \leqq \phi(x)|\xi|$, it has been shown by Ettlinger* that there exists a unique solution of (3.2) satisfying the initial condition $y(0) = a$, where $a$ is an arbitrary vector in Hilbert space. Since $\phi(x)$ is of finite value on $X_0$, the the above condition demands that on $X_0$ the matrix $A(x)$ be limited by the corresponding finite value of $\phi(x)$. Since on $X - X_0$ the function $\phi(x)$ is infinite, we may say that $A(x)$ is limited by $\phi(x)$ on $X$. It then follows that for every vector $\xi$ of Hilbert space, $|\xi A(x)| \leqq \phi(x)|\xi|$.† Under this condition

* See second footnote on p. 286.

† H. T., p. 301.

it follows from the existence theorem of Ettlinger that there exists a unique solution of the differential equation

$$z'(x) = -z(x)A(x) + b(x) \tag{3.3}$$

satisfying the initial condition $z(0)=a$. The homogeneous equation

$$z'(x) = -z(x)A(x) \tag{3.4}$$

is called the adjoint equation of (3.1)

**Definition.** Let $Y(x)\equiv(Y_{ij}(x))$ be a matrix each column of which is a solution of the equation (3.1). Then $Y(x)$ is a *matrix* of (3.1). Similarly, if $Z(x)$ is a matrix each row of which is a solution of (3.4), then $Z(x)$ is a *matrix* of the adjoint equation. If for each value of $x$ on $X$ the matrices $Y(x)$ and $Z(x)$ are limited, then $Y(x)$ and $Z(x)$ are *limited matrices* of (3.1) and (3.4) respectively. In particular, $Y(x)$ is the *principal matrix* of (3.1) at a point $x_0$ of $X$ if $Y(x_0)=E$, the unit matrix. Likewise $Z(x)$ is the *principal matrix* of (3.4) at a point $x_0$ if $Z(x_0)=E$.

Ettlinger first showed the existence of a solution of the associated algebraic system and then that such a solution passes over in the limit to a solution of the differential system. We will here outline a proof of an existence theorem by the method of successive approximations where we do not consider the algebraic system, since the approximations there used are of use to us in considering the properties of the solution. The existence theorem we state as follows:

THEOREM 3.1. *If $B(x)$ is a matrix each element of which is summable on $X$ and the matrix $\int_0^x B(t)dt$ is limited by a finite constant uniformly with respect to $x$ on $X$, and $C$ is a constant limited matrix, then there exists a unique absolutely continuous matrix $Y(x)$ such that*

$$Y(x) = \int_0^x [A(t)Y(t) + B(t)]dt + C, \tag{3.5}$$

*and $Y(x)$ is limited uniformly on $X$ by a finite constant.*

Let

$$Y^{(0)}(x) = \int_0^x B(t)dt + C,$$

$$Y^{(k)}(x) = \int_0^x [A(t)Y^{(k-1)}(t) + B(t)]dt + C \quad (k = 1,2,\cdots), \tag{3.6}$$

$$H^{(0)}(x) = Y^{(0)}(x), \quad H^{(k)}(x) = Y^{(k)}(x) - Y^{(k-1)}(x) \quad (k = 1,2,\cdots).$$

Then

$$Y^{(k)}(x) = \sum_{\alpha=0}^{k} H^{(\alpha)}(x),$$

and

$$H^{(k)}(x) = \int_0^x A(t)H^{(k-1)}(t)dt \qquad (k = 1,2, \cdots). \tag{3.7}$$

In the proof of Theorem 3.1 we make repeated use of the following lemma which we state without proof:

LEMMA 3.1. *If $M(x)$ is a matrix each element of which is summable on $X$ and $M(x)$ is limited on $X$ by the summable function $\Phi(x)$, then for every $x$ on $X$ the matrix $\int_0^x M(t)dt$ is limited by the quantity $\int_0^x \Phi(t)dt$.**

Since the matrix $\int_0^x B(t)dt$ is limited by a finite constant uniformly on $X$, there exists a finite constant $K$ which limits the matrix $\int_0^x B(t)dt+C$ uniformly on $X$. Now by (3.7),

$$H^{(1)}(x) = \int_0^x A(t)H^{(0)}(t)dt,$$

and since $A(x)$ is limited by $\phi(x)$ and $H^{(0)}(x)$ is limited by $K$, we have by Lemma 3.1 that $H^{(1)}(x)$ is limited by $K\int_0^x \phi(t)dt$. By mathematical induction we may prove that $H^{(k)}(x)$ is limited by the quantity

$$K\left[\int_0^x \phi(t)dt\right]^k/k! \qquad (k = 1,2, \cdots).$$

Therefore $\sum_{\alpha=0}^{\infty} H^{(\alpha)}(x) = \lim_{k\to\infty} Y^{(k)}(x)$ exists and is limited by the constant $K \exp\left[\int_0^1 \phi(t)dt\right]$ uniformly on $X$. Let $Y(x)$ denote $\lim_{k\to\infty} Y^{(k)}(x)$. Then since $Y(x) - Y^{(k)}(x) = \sum_{\alpha=1}^{\infty} H^{(k+\alpha)}(x)$ is limited by the constant $\sum_{\alpha=1}^{\infty}\left[\int_0^1 \phi(t)dt\right]^{k+\alpha} \cdot [(k+\alpha)!]^{-1}$, the matrix $\int_0^x A(t)[Y(t) - Y^{(k)}(t)]dt$ is limited for every $x$ by the constant $\sum_{\alpha=1}^{\infty}\left[\int_0^1 \phi(t)dt\right]^{k+1+\alpha}/(k+\alpha)!$ and therefore

$$\lim_{k\to\infty} \int_0^x A(t)Y^{(k)}(t)dt = \int_0^x A(t)Y(t)dt. \tag{3.8}$$

Hence $Y(x)$ is a solution of (3.5).

---

* In particular let $M_{\alpha 1}(x) = f_\alpha(x)$, $M_{ij}(x) = 0$ if $j \neq 1$, where $f(x) \equiv (f_\alpha(x))$ is a vector which is, together with its length, summable on $X$. Then for every $x$ on $X$, $M(x)$ is limited by $|f(x)|$ and by the above lemma $\int_0^x M(t)dt$ is limited by $\int_0^x |f(t)|dt$ and therefore

$$\left(\sum_{\alpha=1}^{\infty}\left[\int_0^x f_\alpha(t)dt\right]^2\right)^{1/2} \equiv \left|\int_0^x f(t)dt\right| \leqq \int_0^x |f(t)|dt.$$

This theorem was used by Ettlinger in the proof of his existence theorem and is stated for the finite case by Bliss, Annals of Mathematics, (2), vol. 6 (1904–1905), p. 59. This theorem will be used frequently in this paper.

Let $B_k(x)$ and $C_k$ denote the vectors $(B_{ak})$ and $(C_{ak})$ respectively and $Y_k^*(x) \equiv (Y_{ak}^*(x))$ be any solution of the vector differential equation

$$y(x) = \int_0^x [A(t)y(t) + B_k(t)]dt + C_k.$$

Then $Y^{**}(x) \equiv (Y_{ij}^{**}(x))$, where $Y_{ij}^{**} = Y_{ij}$ if $j \neq k$ and $Y_{ik}^{**} = Y_{ik}^*$, is a matrix that is a solution of (3.5) and is limited uniformly on $X$ by a finite constant $K^*$. Furthermore the matrix $Y^{(0)}(x) - Y^{**}(x)$ is limited uniformly on $X$ by the constant $K^* + K$ and since

$$Y^{(k)}(x) - Y^{**}(x) = \int_0^x A(t)[Y^{(k-1)}(t) - Y^{**}(t)]dt \qquad (k = 1, 2, \cdots),$$

it follows that for each integer $k$ the matrix $Y^{(k)}(x) - Y^{**}(x)$ is limited by $[K^* + K][\int_0^x \phi(t)dt]^k/k!$ and therefore $Y^{**}(x) \equiv Y(x)$ and the solution of (3.5) is unique.

COROLLARY 1. *There exists a unique solution of* (3.2) *satisfying the initial condition* $y(0) = a$, *where* $a$ *is an arbitrary vector in Hilbert space.*

COROLLARY 2. *If* $y(x)$ *is a solution of* (3.2), *then* $y(x)$ *is a vector which is strongly continuous on* $X$.

COROLLARY 3. *If* $y(x)$ *is a solution of* (3.2), *then* $|y(x)|$ *is an absolutely continuous function on* $X$.

COROLLARY 4. *If* $y(x)$ *is a solution of* (3.2), *then the infinite series* $\sum_{a=1}^{\infty} y_a^2(x)$ *converges uniformly with respect to* $x$ *on* $X$.

COROLLARY 5. *If* $Y(x)$ *is a matrix of the equation* (3.1) *and for some point* $x_0$ *of* $X$ *the matrix* $Y(x_0)$ *is limited by a constant* $l$, *then* $Y(x)$ *is a limited matrix of* (3.1) *and is limited uniformly on* $X$ *by the constant* $l \exp [\int_0^1 \phi(t)dt]$.

If $U(x)$ and $V(x)$ are absolutely continuous matrices which are limited by a finite constant uniformly on $X$ and the matrices $dU/dx$ and $dV/dx$ are limited by a summable function on $X$, it follows from Corollary 2 of Theorem 2.1 that the matrix $U(x)V(x)$ is an absolutely continuous matrix and, on $X_0$,

$$\frac{d(UV)}{dx} = \frac{dU}{dx}V + U\frac{dV}{dx}.$$

From this result we obtain

THEOREM 3.2. *If* $Y(x)$ *and* $Z(x)$ *are limited matrices of* (3.1) *and* (3.4) *respectively, then the matrix* $Z(x)Y(x)$ *is a constant matrix on* $X$.

THEOREM 3.3. *If $Y(x)$ is a limited matrix of* (3.1) *and for some point $x_0$ of $X$ the matrix $Y(x_0)$ possesses a unique limited reciprocal, then there exists a matrix $Z(x)$ which is the unique limited reciprocal of $Y(x)$ on $X$, and $Z(x)$ is a limited matrix of the adjoint equation* (3.4).

By hypothesis there exists a limited matrix $T$ such that $TY(x_0)=Y(x_0)T=E$. By Corollary 5 of Theorem 3.1 there exists a unique limited matrix $Z(x)$ of (3.4) such that $Z(x_0)=T$. Then $Z(x_0)Y(x_0)=E$ and by Theorem 3.2 we obtain $Z(x)Y(x)=E$ for every $x$ on $X$. If for some point $\bar{x}$ of $X$ the matrix $Y(x)$ possessed a left-hand reciprocal $V$ and $V\neq Z(\bar{x})$, by the same argument there would exist a limited matrix $Z^*(x)$ of (3.4) such that $Z^*(\bar{x})=V$ and therefore $Z^*(x)Y(x)=E$ for every $x$ on $X$. Since $Z^*(x_0)=Z(x_0)=T$ this is impossible as $Z(x)$ is the unique limited matrix of (3.4) satisfying the condition $Z(x_0)=T$. Hence $Z(x)$ is the unique left-hand reciprocal of $Y(x)$, and therefore it follows that $Z(x)$ is the unique limited reciprocal of $Y(x)$ on $X$.†

COROLLARY. *If $Y(x)$ and $Z(x)$ are principal matrices of* (3.1) *and* (3.2) *respectively at some point $x=x_0$ of $X$, then $Y(x)$ and $Z(x)$ are unique limited reciprocals on $X$.*

**Definition.** If $Y(x)$ is a limited matrix of (3.1) and possesses a unique limited reciprocal matrix on $X$, then $Y(x)$ is a *matrix solution* of (3.1). A *matrix solution* of the adjoint equation (3.4) is defined in a corresponding manner.

We will state without proof the following theorems:

THEOREM 3.4. *If $Y(x)$ is a limited matrix of the equation* (3.1) *and $C$ is a constant limited matrix, then $Y(x)C$ is a limited matrix of* (3.1).

COROLLARY. *If $Y(x)$ is a matrix solution of* (3.1) *and $C$ is a constant limited matrix which possesses a unique limited reciprocal, then $Y(x)C$ is a matrix solution of* (3.1).

THEOREM 3.5. *If $Y(x)$ is a matrix solution of* (3.1), *then the most general limited matrix of* (3.1) *is of the form $Y(x)C$, where $C$ is a constant limited matrix.*

COROLLARY. *The most general matrix solution of the equation* (3.1) *is of the form $Y(x)C$, where $Y(x)$ is a matrix solution and $C$ is a constant limited matrix which has a unique limited reciprocal.*

THEOREM 3.6. *If $Y^*(x)$ is any matrix solution of* (3.1) *and the matrix $Z(x)$ is defined by the relation $Z(x)Y^*(x)=C$, where $C$ is a constant limited matrix, then $Z(x)$ is a limited matrix of* (3.4).

† H. T., p. 312

COROLLARY. *If the constant limited matrix $C$ has a unique limited reciprocal, then $Z(x)$ is a matrix solution of the adjoint equation* (3.4).

For the adjoint equation we have that the most general limited matrix of (3.4) is of the form $CZ(x)$, where $Z(x)$ is a matrix solution of (3.4) and $C$ is a constant limited matrix. Also every matrix solution of (3.4) may be written in the form $CZ(x)$, where $Z(x)$ is a matrix solution of (3.4) and $C$ is a constant limited matrix which has a unique limited reciprocal.

## 4. INFINITE DIFFERENTIAL SYSTEMS INVOLVING A PARAMETER

**a. Continuity of the solution in $(x, \mu)$.** We will now consider the vector differential equation

$$y'(x;\mu) = A(x;\mu)y(x;\mu), \tag{4.1}$$

which is of the form (3.1) and in which each element of the matrix $A$ is a function of a parameter $\mu$. If for each value of $\mu$ on $L: L_1 < \mu < L_2$ the matrix $A(x;\mu)$ is limited by a summable function on $X: 0 \leqq x \leqq 1$, by Theorem 3.1 there exists for each value of $\mu$ on $L$ a unique solution of (4.1) satisfying the initial condition $y(0;\mu) = a(\mu)$, where $a(\mu)$ is for $\mu$ on $L$ a vector in Hilbert space.

THEOREM 4.1. *Hypotheses*: (a) *There exists a summable function $\phi(x)$ such that $A(x;\mu)$ is limited on $X$ by $\phi(x)$, uniformly with respect to $\mu$ on $L$;* (b) *for values of $x$ on $X_0$ the rows of $A(x;\mu)$ are vectors which are strongly continuous on $L$;* (c) *$a_i(\mu)$ $(i = 1, 2, \cdots)$ is continuous on $L$ and there exists a finite constant $G$ such that $|a(\mu)| < G$ for every $\mu$ on $L$.*

*Conclusion*: *There exists a unique solution of* (4.1) *satisfying the initial condition $y(0;\mu) = a(\mu)$ and $y(x;\mu)$ is continuous in $(x, \mu)$ on $XL$.*

The following lemma will first be established:

LEMMA. *If the matrix $A(x;\mu)$ satisfies hypothesis* (b) *and $v(x;\mu)$ is a vector each component of which is continuous on $L$ for values of $x$ on $X_0$ and there is a finite constant $K$ such that $|v(x;\mu)| < K$ on $XL$, then $A(x;\mu)v(x;\mu)$ is a vector each of whose components is continuous in $\mu$ on $L$ for values of $x$ on $X_0$.*

For every two points $\mu$ and $\mu'$ of $L$ we have

$$\begin{aligned} &\left| \sum_{\alpha=n}^{\infty} [A_{i\alpha}(x;\mu')v_\alpha(x;\mu') - A_{i\alpha}(x;\mu)]v_\alpha(x;\mu) \right| \\ &\leqq \left| \sum_{\alpha=n}^{\infty} [A_{i\alpha}(x;\mu') - A_{i\alpha}(x;\mu)]v_\alpha(x;\mu) \right| \\ &+ \left| \sum_{\alpha=n}^{\infty} A_{i\alpha}(x;\mu')[v_\alpha(x;\mu') - v_\alpha(x;\mu)] \right|. \end{aligned} \tag{4.2}$$

Now by the Lagrange-Cauchy inequality,†

$$(4.3)\qquad \left| \sum_{\alpha=n}^{\infty} [A_{i\alpha}(x;\mu') - A_{i\alpha}(x;\mu)]v_\alpha(x;\mu) \right| \leqq \left[ \sum_{\alpha=n}^{\infty} v_\alpha^2(x;\mu) \right]^{1/2} \left[ \sum_{\alpha=n}^{\infty} [A_{i\alpha}(x;\mu') - A_{i\alpha}(x;\mu)]^2 \right]^{1/2},$$

$$(4.3')\qquad \left| \sum_{\alpha=n}^{\infty} A_{i\alpha}(x;\mu')[v_\alpha(x;\mu') - v_\alpha(x;\mu)] \right| \leqq \left[ \sum_{\alpha=n}^{\infty} A_{i\alpha}^2(x;\mu') \right]^{1/2} \left[ \sum_{\alpha=n}^{\infty} [v_\alpha(x;\mu') - v_\alpha(x;\mu)]^2 \right]^{1/2}.$$

Since each row of $A(x;\mu)$ for $x$ on $X_0$ is strongly continuous on $L$ and since $|v(x;\mu)|$ is bounded uniformly on $XL$, we have for values of $x$ on $X_0$ and for each value $\mu=\mu'$ on $L$ that there exists for every positive $\epsilon$ a positive number $N_{\epsilon x\mu' i}$ and a positive quantity $\delta_{\epsilon x\mu' i}$ such that if $n>N_{\epsilon x\mu' i}$ and $|\mu-\mu'|<\delta_{\epsilon x\mu' i}$ then

$$\left| \sum_{\alpha=n}^{\infty} A_{i\alpha}(x;\mu')v_\alpha(x;\mu') - \sum_{\alpha=n}^{\infty} A_{i\alpha}(x;\mu)v_\alpha(x;\mu) \right| < \epsilon.$$

Therefore, since for all values of $x$ on $X_0$ the finite sum $\sum_{\alpha=1}^{n} A_{i\alpha}(x;\mu)v_\alpha(x;\mu)$ is continuous in $\mu$, it follows that for values of $x$ on $X_0$ the infinite series $\sum_{\alpha=1}^{\infty} A_{i\alpha}(x;\mu)v_\alpha(x;\mu)$ is a continuous function of $\mu$ on $L$.

We now apply the result of the lemma to the approximations $y^{(k)}(x;\mu)$ which are defined by (3.6) and which converge to the solution $y(x;\mu)$. We have

$$(4.4)\qquad \begin{aligned} y_i^{(0)}(x;\mu) &= a_i(\mu), \\ y_i^{(k)}(x;\mu) &= a_i(\mu) + \int_0^x \sum_{\alpha=1}^{\infty} A_{i\alpha}(t;\mu)y_\alpha^{(k-1)}(t;\mu)dt \qquad (k = 1,2,\cdots). \end{aligned}$$

By hypothesis (c), $|a(\mu)|<G$ and therefore $y^{(k)}(x;\mu)$ and $y(x;\mu)$ are absolutely continuous on $X$, uniformly for $\mu$ on $L$, and $|y^{(k)}(x;\mu)| \leqq G \exp[\int_0^1 \phi(t)dt]$. Then if $y^{(k-1)}(x;\mu)$ is for $x$ on $X_0$ a continuous function of $\mu$, it follows from the above lemma that $\sum_{\alpha=1}^{\infty} A_{i\alpha}(x;\mu)y_\alpha^{(k-1)}(x;\mu)$ is for $x$ on $X_0$ a continuous function of $\mu$ on $L$. But $|\sum_{\alpha=1}^{\infty} A_{i\alpha}(x;\mu)y_\alpha^{(k-1)}(x;\mu)|$ is bounded by the summable function $\phi(x)G \exp[\int_0^1 \phi(t)dt]$ and since by hypothesis (c) each component of $a(\mu)$ is continuous on $L$, it follows from Theorem 2.2 that for fixed values of $x$ on $X$ each component of $y^{(1)}(x;\mu)$ is a continuous function of $\mu$. By mathematical induction it can be shown that $y^{(k)}(x;\mu)$

† See F. Riesz, loc. cit., p. 36.

$(k=1, 2, \cdots)$ is a continuous function of $\mu$. Since $|a(\mu)|$ is bounded uniformly on $L$, the sequence of vectors $y^{(k)}(x; \mu)$ converges to the vector $y(x; \mu)$ uniformly on $XL$ and therefore $y(x; \mu)$ for each value of $x$ on $X$ is a continuous function of $\mu$. But the vector $y(x; \mu)$ is absolutely continuous in $x$ on $X$, uniformly for $\mu$ on $L$, since $|a(\mu)|$ is bounded uniformly on $L$. Therefore $y(x; \mu)$ is continuous in $(x, \mu)$ on $XL$. Also by the use of Corollary 1 of Theorem 2.1 we have that $y^{(k)}(x; \mu)$ $(k=1, 2, \cdots)$ is continuous in $\mu$ uniformly with respect to $x$. Then since the sequence $y^{(k)}(x; \mu)$ converges to $y(x; \mu)$ uniformly on $XL$, we have that $y(x; \mu)$ is continuous in $\mu$ uniformly for $x$ on $X$.

COROLLARY. *If for each value $\mu=\mu'$ on $L$ there exists a positive number $d$ such that on $XL_{\mu'}$, where $L_{\mu'}: \mu'-d \leqq \mu \leqq \mu'+d$, the hypotheses of Theorem* 4.1 *are satisfied, then there exists a unique solution of* (4.1) *satisfying the initial condition $y(0; \mu)=a(\mu)$ and which is continuous in $(x, \mu)$ on $XL$.*

**Definition.** An infinite matrix $M(x; \mu)$ is a *matrix of class S* on $XL$ if (a) $M(x; \mu)$ is summable on $X$ for each value of $\mu$ on $L$ and is limited by a summable function $\theta(x)$ on $X$ uniformly for values of $\mu$ on $L$; (b) if $\mu$ and $\mu'$ are any two points on $L$, then $M(x; \mu)-M(x; \mu')$ is limited by a function $C(x; \mu; \mu')$ which is summable on $X$ and such that $\lim_{\mu \to \mu'} C(x; \mu; \mu')=0$ on $X_0$.

We now state the following auxiliary theorem:

THEOREM 4.2. *If on the interval $x_1 \leqq x \leqq x_2$, $(x_1, x_2)$, $u(x)$ is a non-negative continuous function and, for each value of $x$ on $(x_1, x_2)$,*

$$0 \leqq u(x) \leqq \int_{x_1}^{x} [\theta(t)u(t) + B(t)]dt + F(x), \tag{4.5}$$

*where $\theta(x)$ and $B(x)$ are non-negative summable functions on $(x_1, x_2)$ and $F(x)$ is a non-negative function bounded uniformly on $(x_1, x_2)$ by the constant $Q$, then for each value of $x$ on $(x_1, x_2)$,*

$$0 \leqq u(x) \leqq \left[\int_{x_1}^{x_2} B(t)dt + Q\right] \exp\left[\int_{x_1}^{x_2} \theta(t)dt\right]. \tag{4.6}$$

THEOREM 4.3. *If $A(x; \mu)$ is a matrix of class S on $XL$ and $Y(x; \mu)$ is a matrix solution of the equation* (4.1) *which for some point $x=x_0$ of $X$ is a matrix of class S on $XL$, then $Y(x; \mu)$ is a matrix of class S on $XL$.*

Since $Y(x_0; \mu)$ is constant with respect to $x$, if $Y(x_0; \mu)$ is a matrix of class $S$ there exists a finite constant $K$ which limits $Y(x_0; \mu)$ uniformly with respect to $\mu$ on $L$ and therefore $Y(x; \mu)$ is limited uniformly on $XL$ by the

constant $K^* = K \exp\left[\int_0^1 \phi(t)dt\right]$. Also there is a finite quantity $\overline{C}(\mu; \mu')$ which limits $Y(x_0; \mu) - Y(x_0; \mu')$ for each pair of points $\mu$ and $\mu'$ on $L$ and $\lim_{\mu \to \mu'} \overline{C}(\mu; \mu') = 0$. We have

$$Y(x; \mu') - Y(x; \mu) = \int_{x_0}^{x} [A(t; \mu')Y(t; \mu') - A(t; \mu)Y(t; \mu)]dt + Y(x_0; \mu') - Y(x_0; \mu)$$

and for every vector $\xi$ of Hilbert space

$$[Y(x; \mu') - Y(x; \mu)]\xi = \int_{x_0}^{x} \Big[ A(t; \mu')[Y(t; \mu') - Y(t; \mu)]\xi + [A(t; \mu') - A(t; \mu)]Y(t; \mu)\xi \Big] dt + [Y(x_0; \mu') - Y(x_0; \mu)]\xi.$$

Therefore

$$|\,[Y(x; \mu') - Y(x; \mu)]\xi\,| \leqq \int_{x_0}^{x} \Big[ \phi(t)\,|\,[Y(t; \mu') - Y(t; \mu)]\xi\,| + C(t; \mu; \mu')\,|\,Y(t; \mu)\xi\,| \Big] dt + |\,[Y(x_0; \mu') - Y(x_0; \mu)]\xi\,|$$

and by (4.6),

$$|\,[Y(x;\mu') - Y(x;\mu)]\xi\,| \leqq \Big[ \int_0^1 C(t;\mu;\mu')K^* dt + \overline{C}(\mu;\mu') \Big] \exp\Big[ \int_0^1 \phi(t)dt \Big] |\xi|.$$

Hence the matrix $Y(x; \mu)$ is a matrix of class $S$ on $XL$.

COROLLARY 1. *If $A(x; \mu)$ is a matrix of class $S$ on $XL$ and $Y(x; \mu)$ is the principal matrix of* (4.1) *at some point $x_0$ of $X$, then $Y(x; \mu)$ is a matrix of class $S$ on $XL$.*

COROLLARY 2. *If in the differential equation* (4.1) *the matrix $A(x; \mu)$ is a matrix of class $S$ on $XL$ and $a(\mu)$ is a vector which is strongly continuous on $L$, then the solution $y(x; \mu)$ is for each value of $x$ a vector which is strongly continuous on $L$.*

If $a(\mu)$ is strongly continuous, then $|a(\mu)|$ is bounded uniformly on every finite subinterval $L_{\mu'}: \mu' - d \leqq \mu \leqq \mu' + d$ of $L$. Then $Y(0; \mu)$, where $Y_{i1}(x; \mu) = y_i(x; \mu)$, $Y_{ij}(x; \mu) = 0$ if $j > 1$, is of class $S$ on $XL_{\mu'}$ and the corollary follows directly.

An important example of an equation of the form (4.1) in which the matrix $A(x; \mu)$ is a matrix of class $S$ is an equation in which the parameter enters linearly. Such an equation we may write as

$$y'(x; \mu) = [A(x) + \mu B(x)]y(x; \mu), \tag{4.7}$$

where the matrices $A(x)$ and $B(x)$ are summable and limited by a summable function on $X$.

**b. Differentiation of the solution of (4.1) with respect to the parameter $\mu$.** Differentiation of the solution of a finite system of differential equations with respect to a parameter has been considered by Carathéodory.† For an infinite system of ordinary linear differential equations of the first order we have the following theorem.

THEOREM 4.4. *Hypotheses*: (a) *There exists a summable function $\phi(x)$ such that $A(x;\mu)$ is limited on $X$ by $\phi(x)$, uniformly with respect to $\mu$ on $L$*; (b) *$a(\mu)$ is for each value of $\mu$ on $L$ a vector in Hilbert space each of whose components is continuous on $L$*; (c) *for $x$ on $X_0$ and at some point $\mu=\mu_0$ of $L$, $\partial A_{ij}(x;\mu)/\partial\mu$ and $da_i(\mu)/d\mu$ exist and $(da_i(\mu_0)/d\mu)$ is a point in Hilbert space*; (d) *if $y(x;\mu_0)$ is the solution for $\mu=\mu_0$ of (4.1) such that $y(0;\mu_0)=a(\mu_0)$, then*

$$F_i(x;\mu_0) = \sum_{\alpha=1}^{\infty} \frac{\partial}{\partial\mu} A_{i\alpha}(x;\mu_0)\, y_\alpha(x;\mu_0) \qquad (i = 1,2,\cdots)$$

*exists and $F(x;\mu_0)\equiv(F_i(x;\mu_0))$ is a vector which is, together with its length, summable on $X$;*

(e) $$\lim_{\mu\to\mu_0} \frac{A(x;\mu) - A(x;\mu_0)}{\mu-\mu_0} y(x;\mu) = \frac{\partial}{\partial\mu} A(x;\mu_0)\, y(x;\mu_0)$$

*for $x$ on $X_0$ and the absolute continuity of the integral*

$$\int_0^x \left| \frac{A(t;\mu) - A(t;\mu_0)}{\mu-\mu_0} y(t;\mu) - \frac{\partial}{\partial\mu} A(t;\mu_0)\, y(t;\mu_0) \right| dt$$

*on $X$ is uniform with respect to $\mu$ on $L_{\mu_0}$, where $L_{\mu_0}: \mu_0 - d \leqq \mu \leqq \mu_0 + d$, and $d>0$;*

(f) $$\lim_{\mu\to\mu_0} \frac{a(\mu) - a(\mu_0)}{\mu-\mu_0} = \frac{d}{d\mu} a(\mu_0).$$

*Conclusion*: *$\partial y(x;\mu)/\partial\mu$ exists for $\mu=\mu_0$, and*

$$(4.8)\quad \frac{\partial}{\partial\mu} y(x;\mu_0) = \int_0^x \left[ A(t;\mu_0) \frac{\partial}{\partial\mu} y(t;\mu_0) + \frac{\partial}{\partial\mu} A(t;\mu_0)\, y(t;\mu_0) \right] dt + \frac{d}{d\mu} a(\mu_0).$$

For every value of $\mu$ on $L_{\mu_0}$ we have

$$\frac{y(x;\mu) - y(x;\mu_0)}{\mu-\mu_0} = \int_0^x \left[ A(t;\mu_0) \frac{y(t;\mu) - y(t;\mu_0)}{\mu-\mu_0} + \frac{A(t;\mu) - A(t;\mu_0)}{\mu-\mu_0} y(t;\mu) \right] dt + \frac{a(\mu) - a(\mu_0)}{\mu-\mu_0}.$$

† C. Carathéodory, *Vorlesungen über reelle Funktionen*, Berlin, 1918, p. 682.

Now consider the equation

$$u(x) = \int_0^x [A(t; \mu_0)u(t) + F(t; \mu_0)]dt + \frac{d}{d\mu}a(\mu_0), \tag{4.9}$$

where $F(x; \mu_0)$ is defined by hypothesis (d). In view of the above hypotheses the equation (4.9) has a unique solution $u(x)$ and $|u(x)|$ is an absolutely continuous function on $X$. Then we have

$$\frac{y(x; \mu) - y(x; \mu_0)}{\mu - \mu_0} - u(x) = \int_0^x \Big\{ A(t; \mu_0)\left[\frac{y(t; \mu) - y(t; \mu_0)}{\mu - \mu_0} - u(t)\right] + \frac{A(t; \mu) - A(t; \mu_0)}{\mu - \mu_0} y(t; \mu) - \frac{\partial}{\partial \mu} A(t; \mu_0) y(t; \mu_0) \Big\} dt + \frac{a(\mu) - a(\mu_0)}{\mu - \mu_0} - \frac{d}{d\mu}a(\mu_0),$$

and therefore

$$\left|\frac{y(x; \mu) - y(x; \mu_0)}{\mu - \mu_0} - u(x)\right| \leqq \int_0^x \left[\phi(t)\left|\frac{y(t; \mu) - y(t; \mu_0)}{\mu - \mu_0} - u(t)\right| + B(t; \mu; \mu_0)\right] dt + C(\mu; \mu_0),$$

where

$$B(t; \mu; \mu_0) = \left|\frac{A(t; \mu) - A(t; \mu_0)}{\mu - \mu_0} y(t; \mu) - \frac{\partial}{\partial \mu} A(t; \mu_0) y(t; \mu_0)\right|$$

and

$$C(\mu; \mu_0) = \left|\frac{a(\mu) - a(\mu_0)}{\mu - \mu_0} - \frac{d}{d\mu}a(\mu_0)\right|.$$

Since, for each value of $\mu$,

$$\left|\frac{y(x; \mu) - y(x; \mu_0)}{\mu - \mu_0} - u(x)\right|$$

is continuous on $X$, by (4.6) we have

$$\left|\frac{y(x; \mu) - y(x; \mu_0)}{\mu - \mu_0} - u(x)\right| \leqq \left[\int_0^1 B(t; \mu; \mu_0)dt + C(\mu; \mu_0)\right] \exp\left[\int_0^1 \phi(t)dt\right].$$

In view of hypotheses (e) and (f) we have

$$\lim_{\mu \to \mu_0} \frac{y(x; \mu) - y(x; \mu_0)}{\mu - \mu_0} = u(x)$$

and therefore $\partial y(x; \mu_0)/\partial \mu$ is a point in Hilbert space for values of $x$ on $X$, and

$$\frac{\partial}{\partial\mu}y(x;\mu_0) = \int_0^x \left[A(t;\ \mu_0)\frac{\partial}{\partial\mu}y(t;\mu_0) + \frac{\partial}{\partial\mu}A(t;\mu_0)y(t;\mu_0)\right]dt + \frac{d}{d\mu}a(\mu_0).$$

## 5. Boundary conditions for an infinite system of linear differential equations

**a. Vector equations.** We will state some theorems concerning the solution of vector equations in Hilbert space. The corresponding theorems for a finite system of linear equations involving a finite number of unknowns are well known. By a well known theorem† concerning infinite limited matrices, we have

THEOREM 5.1. *A necessary and sufficient condition that a limited matrix $M$ possess a left-[right-]hand limited reciprocal is that there exist a positive quantity $K$ so that for all vectors $\xi$ of Hilbert space we have $K|M\xi| \geqq |\xi|$ [$K|\xi M| \geqq |\xi|$].*

Let $M \equiv (M_{ij})$ and $N \equiv (N_{ij})$ be two infinite limited matrices. Then we will denote by $(M;\ N)$ the infinite matrix

$$(M;\ N) = \left\| \begin{matrix} M_{11} & N_{11} & M_{12} & N_{12} \cdots \\ M_{21} & N_{21} & M_{22} & N_{22} \cdots \\ M_{31} & N_{31} & M_{32} & N_{32} \cdots \\ \cdots\cdots\cdots\cdots\cdots \end{matrix} \right\|.$$

Clearly $(M;\ N)$ is a limited matrix. By $(M\,|N)$ we will denote the limited matrix

$$(M\mid N) = \left\| \begin{matrix} M_{11} & M_{12} & M_{13} & M_{14} \cdots \\ N_{11} & N_{12} & N_{13} & N_{14} \cdots \\ M_{21} & M_{22} & M_{23} & M_{24} \cdots \\ \cdots\cdots\cdots\cdots\cdots \end{matrix} \right\|.$$

**Definition.** Two infinite matrices $M$ and $N$ belong to class $D_1$[$D_2$] if they are limited matrices and there exists a matrix $(M';\ N')$[$(M'|N')$] which has an infinite number of rows and columns and such that: (a) each column [row] of $(M';\ N')$ [$(M'|N')$] is a column [row] of $(M;\ N)$ [$(M|N)$]; (b) $(M';\ N')$ [$(M'|N')$] has a unique limited reciprocal; (c) there exist an infinite number of columns [rows] of $(M;\ N)$ [$(M|N)$] which do not belong to $(M';\ N')$ [$(M'|N')$].

† This theorem follows from a more general theorem proved by Hyslop, *A contribution to the theory of bounded matrices and quadratic forms with an infinity of variables*, Proceedings of the London Mathematical Society, vol. 24 (1925), pp. 264–304. See also Riesz, loc. cit., pp. 85, 86, et seq.

The columns of $(M'; N')$ may be arranged in the sequence in which they appear in $(M; N)$. That is, if the $i$th column of $(M'; N')$ is the $s_i$th column of $(M; N)$, then $s_1 < s_2 < \cdots$. The infinite matrix whose columns are the columns of $(M; N)$ which do not belong to $(M'; N')$ we will denote by $(M''; N'')$. We will also consider the columns of $(M''; N'')$ arranged in the sequence in which they appear in $(M; N)$. Similarly, the rows of $(M' \mid N')$ are arranged in the sequence in which they occur in $(M \mid N)$ and $(M'' \mid N'')$ is the infinite limited matrix formed by arranging in their original order the rows of $(M \mid N)$ which do not belong to $(M' \mid N')$. It is evident that in general the choice of the submatrices $(M'; N')$ and $(M''; N'')$ or $(M' \mid N')$ and $(M'' \mid N'')$ is not unique.

Let $u \equiv (u_\alpha)$ be any vector in Hilbert space. Let $u' \equiv (u'_\alpha)$ be any vector such that $u'_\alpha = u_{g_\alpha}$ $(\alpha = 1, 2, \cdots)$, $g_1 < g_2 < \cdots$, and such that there exist an infinite number of positive integers not belonging to the sequence $g_1, g_2, \cdots$. We will then write $u \equiv (u'; u'')$, where $u''$ is the vector whose $i$th component is the $i$th component of $u$ which does not belong to the sequence $u_{g_1}, u_{g_2}, \cdots$.

We will now prove the following lemma:

LEMMA 5.1. *If the matrices $M$ and $N$ are of class $D_1$, then the matrix $(M; N)$ does not possess a left-hand reciprocal.*

If $u$ is any vector, then the vectors $u'$ and $u''$ can be so determined that $u \equiv (u'; u'')$ and

$$(M; N)u \equiv (M'; N')u' + (M''; N'')u'',$$

where the matrix $(M'; N')$ is an infinite matrix having a unique reciprocal and $(M''; N'')$ is an infinite matrix. Since $(M'; N')$ has a unique reciprocal, for a given vector $u''$ the vector $u'$ may be defined by $u' = -(M'; N')^{-1} \cdot (M''; N'')u''$. Then $(M; N)u = 0$ and the vector $u$ is non-null if the vector $u''$ is non-null. Hence $(M; N)$ does not have a left-hand reciprocal.

In a similar manner we have

LEMMA 5.2. *If the matrices $M$ and $N$ are of class $D_2$, then the matrix $(M \mid N)$ does not possess a right-hand reciprocal.*

THEOREM 5.2. *If $M$ and $N$ are matrices of class $D_1$, there exist matrices $H$ and $K$ which are of class $D_2$ and such that $(M; N)(H \mid K) = 0$.*

We may write

$$(M; N)(H \mid K) \equiv (M'; N')(H' \mid K') + (M''; N'')(H'' \mid K''),$$

where $(M'; N')$ and $(M''; N'')$ are infinite matrices and $(M'; N')$ possesses

a unique limited reciprocal. Let $(H''\,|\,K'')$ be any limited matrix possessing a unique limited reciprocal. Then let

$$(H'\,|\,K') = -(M';\,N')^{-1}(M'';\,N'')(H''\,|\,K'').$$

Then we have $(M;\,N)(H\,|\,K)=0$ and clearly the matrices $H$ and $K$ are of class $D_2$.

**Definition.** If $M$ and $N$ are matrices of class $D_1$ and $H$ and $K$ are matrices of class $D_2$ and such that

$$(M;\,N)(H\,|\,K) \equiv (M';\,N')(H'\,|\,K') + (M'';\,N'')(H''\,|\,K'') = 0,$$

where the matrices $(M';\,N')$, $(M'';\,N'')$, $(H'\,|\,K')$ and $(H''\,|\,K'')$ are infinite matrices and $(M';\,N')$ and $(H''\,|\,K'')$ possess unique limited reciprocals, then the matrix $(H\,|\,K)$ will be said to be the *complete matrix solution* of the vector system

$$(M;\,N)u = 0. \tag{5.1}$$

THEOREM 5.3. *If $(H\,|\,K)$ is a complete matrix solution of the vector system* (5.1), *then every vector solution of this system is linearly dependent† on the columns of $(H\,|\,K)$.*

If $v$ is any vector solution of the system (5.1) we have $(M;\,N)v \equiv (M';\,N')v' + (M'';\,N'')v'' = 0$, where $v \equiv (v';\,v'')$. Now we have

$$v'' = (H''\,|\,K'')(H''\,|\,K'')^{-1}v'',$$

and since each component of $v'$ is a linear expression in the components of $v''$, it follows that $v' = (H'\,|\,K')(H''\,|\,K'')^{-1}v''$. Therefore $v = (H\,|\,K) \cdot (H''\,|\,K'')^{-1}v''$ and $v$ is a vector which is linearly dependent on the columns of $(H\,|\,K)$.

Let $H$, $H^*$, $K$ and $K^*$ be four limited infinite matrices. We define the infinite matrix $(H;\,K\,|\,H^*;\,K^*)$ as follows:

$$(H;\,K\,|\,H^*;\,K^*) = ((H;\,K)\,|\,(H^*;\,K^*)) = ((H\,|\,H^*);\,(K\,|\,K^*))$$

or

† We will say that an infinite sequence of vectors $v^{(1)}, v^{(2)}, \cdots$ are linearly dependent if there exists a non-null vector $\xi$ of Hilbert space such that $\xi_a v_j^{(a)} = 0$ $(j = 1, 2, \cdots)$. If $\xi_r \neq 0$ then $v^{(r)}$ is linearly dependent on the remaining vectors. If a set of vectors are not linearly dependent we will say that they are linearly independent. This is a more general notion of linear dependence for an infinite set of vectors than that used by Kowalewski. See Kowalewski, *Einführung in die Determinantentheorie*, Leipzig, 1909, p. 424.

$$(H;\ K\,|\,H^*;\ K^*) = \left\| \begin{array}{llll} H_{11} & K_{11} & H_{12} & K_{12}\cdots \\ H_{11}^* & K_{11}^* & H_{12}^* & K_{12}^*\cdots \\ H_{21} & K_{21} & H_{22} & K_{22}\cdots \\ H_{21}^* & K_{21}^* & \cdots\cdots\cdots & \\ \cdots\cdots\cdots & & & \end{array} \right\|.$$

Suppose that the matrix $(H;\ K\,|H^*;\ K^*)$ has a unique limited reciprocal $G\equiv(G_{ij})$. Then, considering $G(H;\ K\,|H^*;\ K^*)=E$, we have

$$\begin{aligned} G_{2i-1,2\alpha-1}H_{\alpha j} + G_{2i-1,2\alpha}H_{\alpha j}^* &= E_{ij}, \\ G_{2i,2\alpha-1}H_{\alpha j} + G_{2i,2\alpha}H_{\alpha j}^* &= 0, \\ G_{2i-1,2\alpha-1}K_{\alpha j} + G_{2i-1,2\alpha}K_{\alpha j}^* &= 0, \\ G_{2i,2\alpha-1}K_{\alpha j} + G_{2i,2\alpha}K_{\alpha j}^* &= E_{ij} \qquad (i,j=1,2,\cdots). \end{aligned} \tag{5.2}$$

Also if we consider the product $(H;\ K\,|H^*;\ K^*)G=E$, we have

$$\begin{aligned} H_{i\alpha}G_{2\alpha-1,2j-1} + K_{i\alpha}G_{2\alpha,2j-1} &= E_{ij}, \\ H_{i\alpha}G_{2\alpha-1,2j} + K_{i\alpha}G_{2\alpha,2j} &= 0, \\ H_{i\alpha}^*G_{2\alpha-1,2j-1} + K_{i\alpha}^*G_{2\alpha,2j-1} &= 0, \\ H_{i\alpha}^*G_{2\alpha-1,2j} + K_{i\alpha}^*G_{2\alpha,2j} &= E_{ij} \qquad (i,j=1,2,\cdots). \end{aligned} \tag{5.3}$$

The equations (5.2) and (5.3) give necessary and sufficient conditions that the matrix $G$ be the unique limited reciprocal of the matrix $(H;\ K\,|H^*;\ K^*)$. The conditions may be stated in matrix form as follows:

THEOREM 5.4. *A necessary and sufficient condition that the limited matrix $(H;\ K\,|H^*;\ K^*)$ have a unique limited reciprocal is that there exist four limited matrices $R$, $S$, $R^*$ and $S^*$ satisfying the following matrix equations*:

$$\begin{aligned} R^*H + RH^* &= E, \\ R^*K + RK^* &= 0, \\ S^*H + SH^* &= 0, \\ S^*K + SK^* &= E; \end{aligned} \tag{5.2'}$$

$$\begin{aligned} HR^* + KS^* &= E, \\ HR + KS &= 0, \\ H^*R^* + K^*S^* &= 0, \\ H^*R + K^*S &= E. \end{aligned} \tag{5.3'}$$

*Furthermore, if $G\equiv(G_{ij})$ is the unique limited reciprocal of $(H;\ K\,|H^*;\ K^*)$, then $G\equiv(R^*;\ R\,|S^*;\ S)$.*

THEOREM 5.5. *Let $M$ and $N$ be two matrices of class $D_1$ and let $(M'; N')$ denote an infinite matrix which has a unique limited reciprocal and each of whose columns is a column of $(M; N)$; also, there exists an infinite matrix $(M''; N'')$ whose columns are the columns of $(M; N)$ which do not belong to $(M'; N')$. Let $T$ be a limited infinite matrix such that $T_1$, the infinite matrix whose columns are the columns of $T$ corresponding to the columns of $(M; N)$ which belong to $(M'; N')$, is the zero matrix and $T_2$, the matrix whose columns are the columns of $T$ which do not belong to $T_1$, is a limited matrix which has a unique limited reciprocal. Then the limited matrix $((M'; N'); (M''; N'')\,|T_1; T_2)$ has a unique limited reciprocal.*

It has been shown that a necessary and sufficient condition that $((M'; N'); (M''; N'')\,|T_1; T_2)$ have a unique limited reciprocal is that there exist a limited matrix which satisfies the equations (5.2) and (5.3). Define the matrix $G$ as follows:

$$
\begin{aligned}
(G_{2i-1,2j-1}) &\equiv (M'; N')^{-1}, \\
(G_{2i,2j-1}) &\equiv 0, \\
(G_{2i,2j}) &\equiv T_2^{-1}, \\
(G_{2i-1,2j}) &\equiv -(M'; N')^{-1}(M''; N'')T_2^{-1}.
\end{aligned} \tag{5.4}
$$

It readily follows that the matrix $G$ thus defined satisfies the equations (5.2) and (5.3) and is therefore the unique limited reciprocal of the matrix $((M'; N'); (M''; N'')\,|T_1; T_2)$.

We may now rearrange the columns of $((M'; N'); (M''; N'')\,|T_1; T_2)$ so that we have the matrix $(M; N\,|M^*; N^*)$, where $T \equiv (M^*; N^*)$. Then $(M; N\,|M^*; N^*)$ also has a unique limited reciprocal which is given by the matrix formed by rearranging the rows of $((M'; N'); (M''; N'')\,|T_1; T_2)^{-1}$ in a corresponding manner.

THEOREM 5.6. *If $M$ and $N$ are matrices of class $D_1$ and $(H\,|K)$ is a complete matrix solution of the vector system $(M; N)\,u = 0$, then there exist four limited matrices $M^*$, $N^*$, $H^*$ and $K^*$ such that the matrices $(M; N\,|M^*; N^*)$ and $(H^*; H\,|K^*; K)$ are unique limited reciprocals.*

Since $M$ and $N$ are matrices of class $D_1$ we may write

$$(M; N)(H\,|\,K) \equiv (M'; N')(H'\,|\,K') + (M''; N'')(H''\,|\,K'') = 0,$$

where the matrices $(M'; N')$ and $(H''; K'')$ have unique limited reciprocals. In Theorem 5.5 define $T_2$ as the matrix $(H''\,|K'')^{-1}$. Then in (5.4),

$$
\begin{aligned}
(G_{2i-1,2j}) &= (H'\,|\,K'), \\
(G_{2i,2j}) &= (H''\,|\,K'').
\end{aligned}
$$

Therefore by Theorem 5.5 the matrix $(M; N \mid M^*; N^*)$, where $(M^*; N^*) = T$, has a unique limited reciprocal and $H^*$ and $K^*$ are uniquely determined so that $(H^*; H \mid K^*; K)$ is its reciprocal.

Clearly theorems analogous to Theorems 5.5 and 5.6 are true for matrices of class $D_2$.

b. **Adjoint boundary conditions.** Consider the infinite system of ordinary linear differential equations of the first order as given in §3 by

$$y'(x) = A(x)y(x). \tag{3.1}$$

With this vector differential equation we associate boundary conditions

$$My(0) + Ny(1) = 0, \tag{5.5}$$

where $M$ and $N$ are infinite matrices of class $D_1$. We have that the adjoint system of equations is

$$z'(x) = -z(x)A(x). \tag{3.4}$$

Consider now the vector equation

$$z(0)P + z(1)Q = 0, \tag{5.6}$$

where the matrix $(-P \mid Q)$ is a complete matrix solution of the vector system $(M; N)u = 0$. From Theorem 5.6 and the equations (5.2′) we have that there exist four limited matrices $M^*$, $N^*$, $P^*$ and $Q^*$ such that

$$\begin{aligned} -P^*M - PM^* &= E, \\ Q^*M + QM^* &= 0, \\ -P^*N - PN^* &= 0, \\ Q^*N + QN^* &= E. \end{aligned} \tag{5.7}$$

Let

$$\begin{aligned} s_i(y) &= M_{i\alpha}y_\alpha(0) + N_{i\alpha}y_\alpha(1), \\ \bar{s}_i(y) &= M^*_{i\alpha}y_\alpha(0) + N^*_{i\alpha}y_\alpha(1), \\ t_i(z) &= P_{\alpha i}z_\alpha(0) + Q_{\alpha i}z_\alpha(1), \\ \bar{t}_i(z) &= P^*_{\alpha i}z_\alpha(0) + Q^*_{\alpha i}z_\alpha(1). \end{aligned} \tag{5.8}$$

Then in view of (5.5), (5.6) and (5.7) we have

$$s_\alpha(y)\bar{t}_\alpha(z) + \bar{s}_\alpha(y)t_\alpha(z) = y_\alpha(x)z_\alpha(x)\Big|_{x=0}^{x=1}. \tag{5.9}$$

This equation is an identity for all vectors $y(x)$ and $z(x)$ which satisfy (5.5) and (5.6) respectively. We then say that the set of boundary conditions

(5.6) is adjoint to the set (5.5) and that the differential system (3.4), (5.6) is adjoint to the differential system (3.1), (5.5).†

The matrices $P$ and $Q$ have been so chosen that the matrix $(-P|Q)$ is a complete matrix solution of the system $(M; N)u=0$. There are an infinite number of distinct complete matrix solutions of this system, but according to Theorem 5.3 every solution of this system is linearly dependent on any complete matrix solution. Hence two different choices of the complete matrix solution of $(M; N)u=0$ give sets of equations $z(0)P+z(1)Q=0$ which are equivalent.

If the matrices $M$ and $N$ have unique limited reciprocals $M^{-1}$ and $N^{-1}$, then we may choose $P=M^{-1}$ and $Q=N^{-1}$. Then the adjoint boundary conditions become

$$z(0)M^{-1} + z(1)N^{-1} = 0. \tag{5.10}$$

## 6. A BOUNDARY VALUE PROBLEM FOR AN INFINITE SYSTEM OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS

**a. Completely continuous matrices.** We now consider the bilinear form

$$\xi K\eta \equiv \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} K_{ij}\xi_i\eta_j, \tag{6.1}$$

in which $\xi\equiv(\xi_a)$ and $\eta\equiv(\eta_a)$ are vectors in Hilbert space. For every positive integer $n$, let

$$\xi K_n\eta = \sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}\xi_i\eta_j. \tag{6.2}$$

According to Hilbert‡ we have the following

**Definition.** A bilinear form $\xi K\eta$ is *completely continuous* if for every positive $\epsilon$ there exists a positive $N(\epsilon)$ such that if $\xi$ and $\eta$ are any pair of vectors of Hilbert space whose moduli do not exceed unity, then for all integers $m$ and $n$ such that $m, n>N(\epsilon)$,

$$|\xi K_n\eta - \xi K_m\eta| < \epsilon. \tag{6.3}$$

† This definition of adjoint boundary conditions is equivalent for a finite system to that given by Bounitzky. See Bounitzky, loc. cit., p. 73, and also Bliss, loc. cit., p. 564. We might also define a set of boundary conditions $z(0)H+z(1)K=0$, where $H$ and $K$ are limited matrices, as being adjoint to the set (5.5) if for every vector $z(x)$ satisfying this set and every vector $y(x)$ satisfying (5.5) we have

$$z_a(x)y_a(x)\Big|_{x=0}^{x=1}=0.$$

In a manner analogous to that used by Bounitzky to establish necessary conditions for the adjoint boundary conditions, it can be shown then that there exist matrices $P$ and $Q$ such that $(-P|Q)$ is a complete matrix solution of $(M; N)u=0$ and $z(0)P+z(1)Q=0$. Hence these two definitions are equivalent.

‡ Hilbert, loc. cit., p. 147 and p. 175.

We will say that a matrix $K \equiv (K_{ij})$ is completely continuous if it is the matrix of a bilinear form that is completely continuous. If $K$ is a completely continuous matrix, then clearly $K$ is a limited matrix. In view of Theorem 5.1 and a theorem of Hilbert,† we may state

THEOREM 6.1. *If $K$ is a completely continuous matrix, then a necessary and sufficient condition that the vector equation $(E+K)u=0$ have only the identically vanishing solution is that the matrix $(E+K)$ possess a unique reciprocal.*

Hilbert‡ has also proved the following theorems:

THEOREM A. *If $K$ is a completely continuous matrix and $H$ is a limited matrix, then $KH$ and $HK$ are completely continuous matrices.*

COROLLARY. *If $K$ is a completely continuous matrix such that $(E+K)$ has a unique limited reciprocal $(E+K)^{-1}$, then $(E+K)^{-1}$ is of the form $(E+K^*)$, where $K^*$ is a completely continuous matrix.*

THEOREM B. *If $K$ is a completely continuous matrix, then the vector system $(E+K)u=0$ has at most a finite number of linearly independent solutions.*

THEOREM C. *The number of linearly independent solutions of the system $(E+K)u=0$ is the same as the number of linearly independent solutions of the system $u(E+K)=0$.*

Let $M \equiv (M_{ij})$ be an infinite limited matrix. If $s$ is a positive integer, we will say that the matrix $H \equiv (H_{ij})$ is formed by omitting the $s$th column of $M$ if

$$H_{ij} = M_{ij} \qquad (j < s;\ i = 1,2,\cdots),$$
$$H_{ij} = M_{ij+1} \qquad (j \geqq s;\ i = 1,2,\cdots).$$

In a similar way we define a matrix formed by omitting a row of $M$. This idea may be extended readily to defining a matrix formed by omitting a finite or infinite number of rows and columns of $M$.

THEOREM 6.2. *Let $K$ be a completely continuous matrix. If $(E+K)u=0$ has exactly $r$ linearly independent solutions, $U_1, U_2, \cdots, U_r$, where $U_k \equiv (U_{ak})$, then no matrix with an infinite number of rows and columns and formed from $(E+K)$ by omitting $s$ columns, $s<r$, and by omitting a countable number of rows has a unique limited reciprocal.*

Let $G$ denote the matrix formed by omitting from $(E+K)$ the columns of subscripts $g_1, g_2, \cdots, g_s$, and the rows of subscripts $l_1, l_2, \cdots$, where

† Hilbert, loc. cit., p. 165.

‡ Hilbert, loc. cit., p. 152, p. 165 et seq.

there are an infinite number of positive integers which do not belong to the sequence $l_1, l_2, \cdots$. Since $s<r$ there exist constants $h_1, h_2, \cdots, h_r$ which are not all zero and such that

$$\sum_{\alpha=1}^{r} h_\alpha U_{g_i\alpha} = 0 \qquad (i = 1,2, \cdots, s). \tag{6.4}$$

If the matrix $G$ has a unique limited reciprocal and $u$ is a vector solution of $(E+K)u=0$, then every component of $u$ is expressible linearly in terms of $u_{g_1}, u_{g_2}, \cdots, u_{g_s}$. But $\sum_{\alpha=1}^{r} h_\alpha U_\alpha$ is a vector solution of $(E+K)u=0$ and in view of (6.4) we then have $\sum_{\alpha=1}^{r} h_\alpha U_\alpha = 0$. Since this is impossible if $U_1, U_2, \cdots, U_r$ are linearly independent, we have that the matrix $G$ cannot possess a unique limited reciprocal.

THEOREM 6.3. *If the system $(E+K)u=0$, where $K$ is a completely continuous matrix, has exactly $r$ linearly independent solutions, then there exists a matrix formed by omitting $r$ rows and $r$ columns of $(E+K)$ which possesses a unique limited reciprocal.*

Let $U_1, \cdots, U_r$, where $U_k \equiv (U_{\alpha k})$, denote the $r$ linearly independent solutions of $(E+K)u=0$. Then there exists a matrix

$$\begin{Vmatrix} U_{g_1 1} & U_{g_1 2} \cdots U_{g_1 r} \\ U_{g_2 1} & U_{g_2 2} \cdots U_{g_2 r} \\ \cdots\cdots\cdots\cdots \\ U_{g_r 1} & U_{g_r 2} \cdots U_{g_r r} \end{Vmatrix}$$

whose determinant is different from zero. By Theorem C there are exactly $r$ linearly independent solutions of the vector system $u(E+K)=0$. Let $V_1, V_2, \cdots, V_r$, where $V_k \equiv (V_{k\alpha})$, denote these solutions. Then there is a matrix

$$\begin{Vmatrix} V_{1q_1} & V_{1q_2} \cdots V_{1q_r} \\ V_{2q_1} & V_{2q_2} \cdots V_{2q_r} \\ \cdots\cdots\cdots\cdots \\ V_{rq_1} \qquad\quad \cdots V_{rq_r} \end{Vmatrix}$$

whose determinant is different from zero. Let $G$ denote the submatrix of $(E+K)$ formed by omitting the columns of subscripts $g_1, g_2, \cdots, g_r$ and the rows of subscripts $q_1, q_2, \cdots, q_r$. Suppose the matrix $G$ does not possess a unique limited reciprocal. $G$ is clearly of the form $(E+L)$, where $L$ is a completely continuous matrix. Furthermore, since the determinant $|(V_{iq_j})|$ is different from zero, the rows of subscripts $q_1, q_2, \cdots, q_r$ are linearly dependent on the remaining rows of $(E+K)$. Then by Theorem 6.1 there

would exist a non-null vector $v$ such that $Gv=0$ and therefore there is a non-null vector $u$ such that $u_{g_\alpha}$ $(\alpha=1, 2, \cdots, r)$ is zero and $(E+K)u=0$. But this is impossible since $u$ must be by hypothesis linearly dependent on $U_1, U_2, \cdots, U_r$. Hence the matrix $G$ has a unique limited reciprocal.

**b. A boundary value problem.** We now consider a class of matrices which are functions of the variable $x$ on $X: 0 \leqq x \leqq 1$.

**Definition.** A matrix $H(x) \equiv (H_{ij}(x))$ is a matrix of class $C^*$ if (a) $H_{ij}(x)$ $(i, j=1, 2, \cdots)$ is a Lebesgue summable function on $X$ and $H(x)$ is limited by a non-negative summable function $\phi(x)$ on $X$; (b) for values of of $x$ on $X_0$, the matrix $H(x)$ is a completely continuous matrix.

THEOREM 6.4. *If $H(x)$ is a matrix of class $C^*$, then the matrix $\int_0^x H(t)dt$ is completely continuous uniformly on $X$.*

For any $\epsilon>0$, let $Q_n(\epsilon)$ $(n=1, 2, \cdots)$ denote the set of all points $x$ on $X$ such that $|\xi H_n(x)\eta - \xi H(x)\eta| > \epsilon$ for some pair of vectors $\xi$ and $\eta$ whose moduli are not greater than unity. Since for values of $x$ on $X_0$ the matrix $H(x)$ is completely continuous, for every $\epsilon>0$ and every $\delta>0$ there exists a positive $N(\epsilon; \delta)$ such that if $n>N(\epsilon; \delta)$, then the interior measure of $Q_n(\epsilon)$ is less than $\delta$.† Therefore, since for every pair of vectors $\xi$ and $\eta$ of moduli not greater than unity the functions $\xi H_n(x)\eta$ and $\xi H(x)\eta$ are measurable on $X$ and $|\xi H_n(x)\eta| \leqq \phi(x)|\xi|\ |\eta| \leqq \phi(x)$, it follows that for every $\epsilon>0$ there is an $N_\epsilon>0$ such that if $n, m>N_\epsilon$ then

$$\left| \xi \int_0^x H_n(t)dt\eta - \xi \int_0^x H_m(t)dt\eta \right| \leqq \int_0^x |\, \xi H_n(t)\eta - \xi H_m(t)\eta \,|\, dt < \epsilon$$

independent of the value of $x$ on $X$. Hence $\int_0^x H(t)dt$ is a completely continuous matrix uniformly on $X$.

COROLLARY. *If $H(x)$ and $K(x)$ are matrices of class $C^*$ which are limited by the summable functions $\phi(x)$ and $\theta(x)$ respectively and $\phi(x)\theta(x)$ is a summable function, then $\int_0^x H(t)K(t)dt$ is a completely continuous matrix uniformly on $X$.*

We now consider the infinite homogeneous differential vector equation

$$y'(x) = A(x)y(x), \tag{6.5}$$

in which the matrix $A(x)$ is of class $C^*$ on $X: 0 \leqq x \leqq 1$. Such a system is a special form of the system (3.1) treated in §3. Then we have

THEOREM 6.5. *If $Y(x)$ is the principal matrix of (6.5) at the point $x=0$, then $Y(x)$ is of the form $E+R(x)$, where $R(x)$ is a matrix which is completely continuous uniformly on $X$.*

† See de la Vallée Poussin, *Cours d'Analyse*, Paris, 1909, vol. 1, p. 252.

With (6.5) we consider the boundary conditions

$$(E + G_1)y(0) + (E + G_2)y(1) = 0, \tag{6.6}$$

where $G_1$ and $G_2$ are constant matrices which are completely continuous and $(E+G_1)$ and $(E+G_2)$ have unique limited reciprocals $(E+T_1)$ and $(E+T_2)$ respectively. That $T_1$ and $T_2$ are completely continuous follows from the corollary to Theorem A. The adjoint differential system is defined as

$$z'(x) = -z(x)A(x), \tag{6.7}$$

$$z(0)(E + T_1) + z(1)(E + T_2) = 0. \tag{6.8}$$

Clearly every finite system of ordinary linear differential equations of the first order in which the elements of the matrix $A(x)$ are summable on $X$ and the finite matrices $(E+G_1)$ and $(E+G_2)$ have non-vanishing determinants is a special case of the system considered.

Since the general solution of the equation (6.5) is of the form $Y(x)c$, where $Y(x)$ is the principal matrix at $x=0$ of (6.5) and $c$ is a constant vector in Hilbert space, if the system (6.5), (6.6) is compatible it is necessary and sufficient that there exist a non-null vector $c$ so that

$$[(E + G_1)Y(0) + (E + G_2)Y(1)]c = 0.$$

Since $[G_1+G_2+R(1)+G_2R(1)]$ is a completely continuous matrix, we have

THEOREM 6.6. *A necessary and sufficient condition that* (6.5), (6.6) *be incompatible is that the matrix* $(E+Q)$, *where* $Q=[G_1+G_2+R(1)+G_2R(1)]/2$, *have a unique limited reciprocal.*

In view of Theorem B we have

THEOREM 6.7. *The system* (6.5), (6.6) *has at most a finite number of linearly independent solutions.*

The corresponding non-homogeneous equation we write as

$$y'(x) = A(x)y(x) + b(x), \tag{6.9}$$

$$(E + G_1)y(0) + (E + G_2)y(1) = h, \tag{6.10}$$

where $A(x)$ is a matrix of class $C^*$, $b(x)$ is a vector which is, together with its length, summable on $X$, and $h \equiv (h_\alpha)$ is a constant vector in Hilbert space.

The following theorems are well known in the finite case and may be proved for the infinite system considered in a manner analogous to that used in the finite case.†

† See, for example, Bliss, loc. cit., p. 566.

THEOREM 6.8. *A necessary and sufficient condition that the non-homogeneous system* (6.9), (6.10) *have a unique solution is that* (6.5), (6.6) *be incompatible.*

THEOREM 6.9. *A necessary and sufficient condition that the adjoint system* (6.7), (6.8) *be compatible is that* (6.5), (6.6) *be compatible.*

THEOREM 6.10. *The number of linearly independent solutions of the adjoint system* (6.7), (6.8) *is the same as the number of linearly independent solutions of the system* (6.5), (6.6).

We now consider a system of the form (6.5), (6.6) which involves a parameter $\mu$. Let $A(x;\mu)$ be an infinite limited matrix which for all values of $\mu$ on $L: L_1 < \mu < L_2$ is a matrix of class $C^*$ on $X: 0 \leqq x \leqq 1$; furthermore, let $A(x;\mu)$ be a matrix of class $S$ on $XL$ as defined in §4. We then have that $A(x;\mu)$ is summable on $X$ and limited by a summable function $\phi(x)$ on $X$, uniformly for $\mu$ on $L$. Also $A(x;\mu)-A(x;\mu')$, for every pair of points $\mu$ and $\mu'$ on $L$, is limited by a summable function $C(x;\mu;\mu')$ which is such that $\lim_{\mu\to\mu'} C(x;\mu;\mu')=0$ on $X_0$. Let $G_1(\mu)$ and $G_2(\mu)$ be matrices which for each value of $\mu$ on $L$ are completely continuous and such that for every pair of points $\mu$ and $\mu'$ on $L$ the matrices $G_1(\mu)-G_1(\mu')$ and $G_2(\mu)-G_2(\mu')$ are limited by finite quantities $C_1(\mu;\mu')$ and $C_2(\mu;\mu')$ respectively such that $\lim_{\mu\to\mu'} C_1(\mu;\mu') = 0 = \lim_{\mu\to\mu'} C_2(\mu;\mu')$. Furthermore $(E+G_1(\mu))$ and $(E+G_2(\mu))$ for each value of $\mu$ on $L$ possess unique limited reciprocals $(E+T_1(\mu))$ and $(E+T_2(\mu))$ respectively. We write this system as

$$y'(x;\mu) = A(x;\mu)y(x;\mu), \tag{6.11}$$

$$(E+G_1(\mu))y(0;\mu) + (E+G_2(\mu))y(1;\mu) = 0. \tag{6.12}$$

Clearly every finite system of ordinary linear differential equations of the first order in which the matrix $A(x;\mu)$ is continuous in $\mu$ on $L$ for values of $x$ on $X_0$ and in which the matrices $E+G_1(\mu)$ and $E+G_2(\mu)$ are continuous in $\mu$ and have non-vanishing determinants is of the form (6.11), (6.12).

The system adjoint to (6.11), (6.12) we write

$$z'(x;\mu) = -z(x;\mu)A(x;\mu), \tag{6.13}$$

$$z(0;\mu)(E+T_1(\mu)) + z(1;\mu)(E+T_2(\mu)) = 0. \tag{6.14}$$

In view of Theorem 6.10 we have

THEOREM 6.11. *If $\mu=\mu'$ is a characteristic number of the system* (6.11), (6.12), *its index, i.e., the number of linearly independent solutions corresponding to it, is finite and $\mu=\mu'$ is also a characteristic number of the adjoint system* (6.13), (6.14) *of the same index.*

For such a system we have

THEOREM 6.12. *If for $\mu=\mu'$ the system* (6.11), (6.12) *has r-fold compatibility, then there exists a positive number $\delta$ such that for all values of $\mu$ on $|\mu-\mu'|\leqq\delta$ the system has at most r-fold compatibility.*

Let $Y(x;\mu)$ be the principal matrix of (6.11) at $x=0$. By Corollary 1 of Theorem 4.3 the matrix $Y(x;\mu)$ is of class $S$ on $XL$. Since $A(x;\mu)$ is for each value of $\mu$ a matrix of class $C^*$, by Theorem 6.5 we have $Y(x;\mu)=E+R(x;\mu)$, where $R(x;\mu)$ is completely continuous for values on $XL$. From Theorem 6.6 we have that a necessary and sufficient condition that the system (6.11), (6.12) be incompatible is that $E+Q(\mu)$, where $Q(\mu)=[G_1(\mu)+G_2(\mu)+R(1;\mu)+G_2(\mu)R(1;\mu)]/2$, have a unique limited reciprocal. Since $R(1;\mu)$ is of class $S$, the matrix $Q(\mu)-Q(\mu')$ is limited by a finite quantity $I(\mu;\mu')$ for each pair of points $\mu$ and $\mu'$ on $L$ and $\lim_{\mu\to\mu'}I(\mu;\mu')=0$. By hypothesis the system (6.11), (6.12) has $r$-fold compatibility for $\mu=\mu'$. By Theorem 6.3 there exists then a submatrix $U(\mu)$ formed by omitting $r$ columns and $r$ rows of $E+Q(\mu)$ which has a unique limited reciprocal for $\mu=\mu'$. Since $U(\mu)-U(\mu')$ is also limited by $I(\mu;\mu')$ it follows in view of Theorem 5.1 that there exists a $\delta$ such that if $|\mu-\mu'|\leqq\delta$, then $U(\mu)$ has a unique limited reciprocal. From Theorem 6.2 we then have that for $\mu$ on $|\mu-\mu'|\leqq\delta$ the system (6.11), (6.12) has at most $r$ linearly independent solutions.

If we consider the case when $r=0$, we have

COROLLARY. *The set of values of $\mu$ which are characteristic numbers of the system* (6.11), (6.12) *is a closed set of points on L.*

THEOREM 6.13. *If $\mu'$ is a point on L and on the interval $|\mu-\mu'|\leqq\delta$ the system* (6.11), (6.12) *has exactly r-fold compatibility, then there exists a solution $y(x;\mu)$ of this system such that* $\mathrm{LIM}_{\mu\to\mu'}\, y(x;\mu)=y(x;\mu')$.

Let $U(\mu)$ be a matrix formed by omitting from $E+Q(\mu)$ $r$ rows of subscripts $q_1, q_2, \cdots, q_r$ and $r$ columns of subscripts $g_1, g_2, \cdots, g_r$ and such that $U(\mu')$ has a unique limited reciprocal. It was shown in the proof of Theorem 6.12 that there exists an interval $|\mu-\mu'|\leqq\delta^*$ $(\delta^*\leqq\delta)$, on which $U(\mu)$ has a unique limited reciprocal. Let $D(\mu)$ denote the matrix formed by omitting the rows of subscripts $q_1, q_2, \cdots, q_r$ of $E+Q(\mu)$. Since by hypothesis the system (6.11), (6.12) has $r$-fold compatibility on $|\mu-\mu'|\leqq\delta^*$, the rows of subscripts $q_1, q_2, \cdots, q_r$ are linearly dependent on the rows of $D(\mu)$. Since $D(\mu)-D(\mu')$ is limited by a quantity which converges to zero as $\mu$ approaches $\mu'$, there exists a vector $c(\mu)$ such that $c(\mu)$ is non-null and $D(\mu)c(\mu)=0$ for $\mu$ on $|\mu-\mu'|\leqq\delta^*$; furthermore, $c(\mu)$ is strongly continuous at $\mu=\mu'$. Then $[E+Q(\mu)]c(\mu)=0$ on $|\mu-\mu'|\leqq\delta^*$. Let $y(x;\mu)=Y(x;\mu)c(\mu)$, where $Y(x;\mu)$ is the principal matrix of (6.11), (6.12) at $x=0$. Since $Y(x;\mu)$ is limited

uniformly on $XL$ by a finite constant and since by Corollary 1 of Theorem 4.3 the matrix $Y(x;\mu)$ is of class $S$, we have that $\mathrm{LIM}_{\mu\to\mu'}\, y(x;\mu)=y(x;\mu')$.

## 7. The Green's matrix for an infinite system of differential equations

Consider again the infinite differential system

$$y'(x) = A(x)y(x), \tag{7.1}$$

$$(E+G_1)y(0) + (E+G_2)y(1) = 0, \tag{7.2}$$

where $A(x)$ is a summable matrix of class $C^*$ limited by the summable function $\phi(x)$ on $X: 0\leqq x\leqq 1$, and $G_1$ and $G_2$ are constant infinite matrices which are completely continuous and $(E+G_1)$ and $(E+G_2)$ have unique limited reciprocals. The properties of this system have been considered in §6.

A matrix $G(x;t)\equiv(G_{ij}(x;t))$ which is defined on $0\leqq x\leqq 1$, $0\leqq t\leqq 1$, is said to be a Green's matrix of the system (7.1), (7.2) if it has the following properties:

(a) if $t$ is a point on $0<t<1$, then $G(x;t)$ is a limited matrix, as defined in §3, of the equation (7.1) on $0\leqq x<t$ and $t<x\leqq 1$;

(7.3)

(b) $$G(t+0;t) - G(t-0;t) = E;$$

(c) $$(E+G_1)G(0;t) + (E+G_2)G(1;t) = 0.$$

Suppose that the system (7.1), (7.2) is incompatible. By the same type of proof that is used to establish the existence of a Green's matrix for a finite system, we may show that a unique Green's matrix exists for the incompatible system (7.1), (7.2) and is of the form

$$G(x;t) = (1/2)Y(x)\left[\frac{|x-t|}{x-t}E + DW\right]Z(t),\dagger \tag{7.4}$$

where $Y(x)Z(x)=E=Z(x)Y(x)$, $D=[MY(0)+NY(1)]^{-1}$, and $W=MY(0)-NY(1)$. In view of the theorems that are proved in §3, it is clear that the matrix $G(x;t)$ is limited by a finite constant uniformly with respect to $x$ and $t$.

The system adjoint to (7.1), (7.2) is

$$z'(x) = -z(x)A(x), \tag{7.5}$$

$$z(0)(E+T_1) + z(1)(E+T_2) = 0, \tag{7.6}$$

where $(E+T_1)$ and $(E+T_2)$ are the unique limited reciprocals of $(E+G_1$

† See Bliss, loc. cit., p. 578.

and $(E+G_2)$ respectively. We have shown in §6 that $T_1$ and $T_2$ are matrices which are completely continuous.

As for a finite system we may prove†

THEOREM 7.1. *If there exists a Green's matrix for the system* (7.1), (7.2), *then the adjoint system* (7.5), (7.6) *is incompatible.*

THEOREM 7.2. *The functions* $H_{ik}(x; t) = -G_{ki}(t; x)$, *where* $G(x; t)$ *is the Green's matrix for the system* (7.1), (7.2), *are the elements of the Green's matrix for the adjoint system* (7.5), (7.6) *such that*

(a) *if* $x$ *is a point on* $0<x<1$, *then* $G(x; t)$ *as a function of* $t$ *is a limited matrix of* (7.5) *on* $0 \leqq t<x$ *and* $x<t \leqq 1$;

(7.7)

(b) $$G(x; x-0) - G(x; x+0) = E;$$

(c) $$G(x; 0)(E + T_1) + G(x; 1)(E + T_2) = 0.$$

Consider the non-homogeneous system

$$y'(x) = A(x)y(x) + b(x), \tag{7.8}$$

where $A(x)$ is a matrix which is limited by a summable function $\phi(x)$ on $X$ and $b(x)$ is a vector which is, together with its length, summable on $X$. With this equation associate boundary conditions of the form (7.2).

Making use of the explicit form of $G(x; t)$ we may prove, as for a finite system,

THEOREM 7.3. *Every solution of the integral equation*

$$y(x) = \int_0^1 G(x; t)b(t)dt, \tag{7.9}$$

*where* $G(x; t)$ *is defined by* (7.4), *is also a solution of the system* (7.8), (7.2) *and conversely every solution of the system* (7.8), (7.2) *can be expressed in the form* (7.9).

We consider again the system

$$y'(x; \mu) = A(x; \mu)y(x; \mu), \tag{6.11}$$

$$(E + G_1(\mu))y(0; \mu) + (E + G_2(\mu))y(1; \mu) = 0, \tag{6.12}$$

where $A(x; \mu)$ is a matrix of class $C^*$ and also a matrix of class $S$ on $XL$. For each value of $\mu$ on $L$, $G_1(\mu)$ and $G_2(\mu)$ are completely continuous matrices such that $E+G_1(\mu)$ and $E+G_2(\mu)$ have unique limited reciprocals $E+T_1(\mu)$ and $E+T_2(\mu)$ respectively; furthermore, for each pair of points $\mu$ and $\mu'$

† Bounitzky, loc. cit., p. 77.

on $L$ the matrices $G_1(\mu)-G_1(\mu')$ and $G_2(\mu)-G_2(\mu')$ are limited by finite quantities $C_1(\mu;\mu')$ and $C_2(\mu;\mu')$ respectively such that $\lim_{\mu\to\mu'} C_1(\mu;\mu')=0$ $=\lim_{\mu\to\mu'} C_2(\mu;\mu')$. Clearly for each value of $\mu$ for which the system (6.11), (6.12) is incompatible we may define uniquely a Green's matrix $G(x;t;\mu)$ in the same manner in which we defined a Green's matrix for the system (7.1), (7.2).

W. M. Whyburn† has considered the defining of a Green's matrix for a finite system of ordinary linear differential equations of the first order for values of the parameter which are characteristic numbers of the system. He showed how to construct a reduced system which is incompatible for the characteristic numbers of the original system and such that the coefficients of the reduced system differ by arbitrarily small amounts from the coefficients of the original system. The theorem as proved by Whyburn applies only to finite systems in which no component of the solution can be identically zero on $X$. We will here prove a related theorem for the infinite system (6.11), (6.12) which is more general in the finite case than Whyburn's theorem.

THEOREM 7.4. *If for $\mu=\mu'$ the system* (6.11), (6.12) *has r-fold compatibility, then there exists a reduced system* (6.11′), (6.12′) *which is incompatible, where*

$$y'(x) = [A(x;\mu') + P(x)]y(x), \tag{6.11'}$$

$$(E + G_1(\mu'))y(0) + (E + G_2(\mu'))y(1) = 0, \tag{6.12'}$$

*and $P(x)$ is a matrix of non-negative continuous functions on $X$ which is limited by an arbitrarily assigned $\epsilon$ and at most $r$ of the elements of $P(x)$ are not identically zero on $X$.*

In the proof of this theorem we will establish the following lemmas:

LEMMA 1. *If $y^*(x)$ is a solution of the system* (6.11′), (6.12′) *and $z(x;\mu')$ is a solution of the system* (6.13), (6.14) *for $\mu=\mu'$, then*

$$\int_0^1 z(x;\mu')P(x)y^*(x)dx = 0.$$

This follows immediately from the equation (5.9) since by hypothesis $y^*(x)$ and $z(x;\mu')$ satisfy the boundary conditions (6.12′) and (6.14) respectively.

† W. M. Whyburn, *On the Green's function for systems of differential equations*, Annals of Mathematics, (2), vol. 28 (1927), pp. 291–300.

LEMMA 2. *If for $\mu=\mu'$ the system* (6.11), (6.12) *has $r$-fold compatibility, then there exists a matrix $P(x)$ all of whose elements except one are identically zero and this element is a non-negative continuous function on $X$ whose absolute value is arbitrarily small and such that the system* (6.11′), (6.12′) *has at most $(r-1)$-fold compatibility.*

Since for $\mu=\mu'$ the system (6.11), (6.12) has $r$-fold compatibility, there exists some component $y_s(x;\mu')$ of a solution $y(x;\mu')$ which is not identically zero on $X$. Let $x_0$ be a point of $X$ so that $y_s(x_0;\mu')\neq 0$. Also since for $\mu=\mu'$ the adjoint system (6.13), (6.14) has $r$-fold compatibility, there is some $k$ so that $z_k(x_0;\mu')\neq 0$. Since $y_s(x;\mu')$ and $z_k(x;\mu')$ are continuous there is an interval $|x-x_0|\leqq d$ throughout which neither $y_s(x;\mu')$ nor $z_k(x;\mu')$ vanishes. Now define a matrix $P^*(x)$ so that $P_{ks}^*(x)$ is positive and continuous on $x_0-d<x<x_0+d$ and equal to zero elsewhere on $X$. We define $P_{ij}^*(x)$, $i\neq k$, $j\neq s$, as zero on $X$. Then

$$\int_0^1 z(x;\mu')P^*(x)y(x;\mu')dx = \int_0^1 z_k(x;\mu')P_{ks}^*(x)y_s(x;\mu')dx \neq 0.$$

Now consider the system

$$y'(x;\mu';\theta) = [A(x;\mu') + \theta P^*(x)]y(x;\mu';\theta), \tag{7.10}$$

with the boundary conditions (6.12′). For $\theta=0$ the system (7.10), (6.12′) has $r$-fold compatibility. The matrix $A(x;\mu')+\theta P^*(x)$, involving the parameter $\theta$, is a matrix of class $S$ for $x$ on $X$ and values of $\theta$ on $0\leqq\theta\leqq 1$. Also, since by hypothesis the matrix $A(x;\mu')$ is of class $C^*$ and since only one element of $P^*(x)$ is not identically zero, the matrix $A(x;\mu')+\theta P^*(x)$ for each value of $\theta$ is of class $C^*$. Furthermore, since the boundary conditions do not involve $\theta$, the system (7.10), (6.12′) is a system of the type (6.11), (6.12). Therefore there exists a positive $d$ such that for $0\leqq\theta\leqq d$ the system (7.10), (6.12′) has at most $r$-fold compatibility. Suppose that for all values of the parameter $\theta$ on $0\leqq\theta\leqq d$ the system (7.10), (6.12′) has exactly $r$-fold compatibility. Then as in Theorem (6.13) there would exist a solution $y(x;\mu';\theta)$, $0\leqq\theta\leqq d$, and so that $\mathrm{LIM}_{\theta=0}\, y(x;\mu';\theta)=y(x;\mu';0)$ and $y(x;\mu';0)=y(x;\mu')$, the solution of (6.11), (6.12) for $\mu=\mu'$. From Lemma 1 we have that, for all values of $\theta$ on this interval,

$$\int_0^1 z(x;\mu')P^*(x)y(x;\mu';\theta)dx = \int_0^1 z_k(x;\mu')P_{ks}^*(x)y_s(x;\mu';\theta)dx = 0.$$

But this is impossible since

$$\lim_{\theta \to 0} \int_0^1 z_k(x\,;\,\mu')\,P^*_{ks}(x)\,y_s(x\,;\,\mu';\theta)\,dx = \int_0^1 z_k(x\,;\,\mu')\,P^*_{ks}(x)\,y_s(x\,;\,\mu')\,dx \neq 0.$$

Therefore there exists a number $\theta_1$, $0<\theta_1 \leqq d$, such that the system (7.10), (6.12′) for $\theta=\theta_1$ has at most $(r-1)$-fold compatibility. Let $P(x)=\theta_1 P^*(x)$. For this value of the matrix $P(x)$ the system (6.11′), (6.12′) has at most $(r-1)$-fold compatibility.

If after choosing a reduced system as above described the reduced system is still compatible, we may choose a reduced system relative to the one already chosen and such that its degree of compatibility will be at least one lower than that of the previous system. Then after repeating this process at most $r$ times we obtain a reduced system which is incompatible. Clearly the elements of the matrix $P(x)$ can be chosen so that $P(x)$ is limited by an arbitrary number $\epsilon$, since we will need to make at most $r$ of the elements of $P(x)$ different from zero.

Now consider a differential system of the form (6.11), (6.12) which has only a countable number of characteristic numbers $\mu_1, \mu_2, \cdots$ on $L$: $L_1<\mu<L_2$ and this set is such that $\mu_1<\mu_2<\mu_3<\cdots$ and the set has only one limit point, $L_2$. For every $\mu_\alpha$ we may define a matrix $P(x;\mu_\alpha)$ satisfying the conditions of Theorem 7.4 and such that the corresponding reduced system of the form (6.11′), (6.12′) is incompatible. A matrix $P(x;\mu)$ may then be defined as follows:

$$P(x;\mu) = P(x;\mu_\alpha) + \frac{\mu-\mu_\alpha}{\mu_{\alpha+1}-\mu_\alpha}[P(x;\mu_{\alpha+1}) - P(x;\mu_\alpha)] \quad \text{on } \mu_\alpha \leqq \mu \leqq \mu_{\alpha+1},$$

$$P(x;\mu) = P(x;\mu_1) \quad \text{on } L_1 < \mu \leqq \mu_1.$$

Then the system

$$y'(x;\mu) = [A(x;\mu) + P(x;\mu)]y(x;\mu), \tag{7.11}$$

together with the boundary conditions (6.12), has no characteristic numbers in common with the system (6.11), (6.12). Then for each characteristic number $\mu_\alpha$ of the system (6.11), (6.12) we may define a Green's matrix for the reduced system (7.11), (6.12). Since the quantity $\epsilon$ by which the matrix $P(x;\mu)$ is limited uniformly on $XL$ may be taken arbitrarily small, we will speak of the Green's matrix for the system (7.11), (6.12) for $\mu=\mu_\alpha$ $(\alpha=1, 2, \cdots)$ as the Green's matrix for the characteristic numbers of the system (6.11), (6.12).

In order to define a Green's matrix for an infinite differential system which contains a parameter, it is not necessary that the system be of the form (6.11), (6.12). However, for a system of the form (6.11), (6.12) the conditions

that $A(x;\ \mu)$ be of class $S$ and that $G_1(\mu)-G_1(\mu')$ and $G_2(\mu)-G_2(\mu')$ be limited by quanitites that converge to zero as $\mu$ approaches $\mu'$ are useful in considering the properties of the Green's matrix as a function of $\mu$.

Consider now the finite linear differential system

$$y_i'(x;\mu) = \sum_{\alpha=1}^{n} A_{i\alpha}(x;\mu)y_\alpha(x;\mu) \quad (i = 1,2,\cdots,n), \tag{7.12}$$

with the boundary conditions

$$\sum_{\alpha=1}^{n} M_{i\alpha}(\mu)y_\alpha(0;\mu) + \sum_{\alpha=1}^{n} N_{i\alpha}(\mu)y_\alpha(1;\mu) = 0 \ (i = 1,2,\cdots,n), \tag{7.13}$$

where we suppose that each element $A_{ij}(x;\mu)$ is summable on $X$ and for values of $x$ on $X_0$ is a continuous function of $\mu$ on $L$. Furthermore, for each value of $\mu$ on $L$ the conditions of (7.13) are linearly independent and $M_{ij}(\mu)$ and $N_{ij}(\mu)$ $(i, j=1, 2,\cdots, n)$ are continuous functions on $L$. If the matrices $M(\mu)\equiv(M_{ij}(\mu))$ and $N(\mu)\equiv(N_{ij}(\mu))$ have non-vanishing determinants, then the finite system (7.12), (7.13) corresponds to the infinite system (6.11), (6.12) that we have considered. But in the more general case where the $n$ boundary conditions of (7.13) are merely linearly independent it is possible to prove for the system (7.12), (7.13) theorems that are analogous to Theorems 6.12 and 6.13.† For the system (7.12), (7.13) we can then prove a theorem which is analogous to Theorem 7.4.

† Whyburn proves these theorems for a finite differential system by a manner analogous to that used in the proof of Theorems 6.12 and 6.13.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

# CONSTRUCTION OF DIVISION ALGEBRAS*

BY

L. E. DICKSON

1. **Introduction.** The main outstanding problem in the theory of linear associative algebras is the determination of all division algebras. We shall make a noteworthy simplification of the theory of the construction of a type of algebras $\Gamma$ which includes all known division algebras. The simplification is so great that it would require a hundred pages to obtain our results by the best earlier method.

The paper gives an elementary exposition from first principles of the simplified construction of algebras $\Gamma$. It can be easily read by those familiar with the concept of real quaternions with the basis $1, i, j, k=ij$, complex quaternions, and, in general, quaternions over any field $F$. The few further terms used will be defined. The paper is independent of earlier literature except that, after proving a result for 3-rowed matrices, we give references to the similar proof for $p$-rowed matrices.

The emphasis on simplicity of exposition is warranted by the importance of the subject and the fact that the theory has now reached its last stage of simplicity. The paper will serve as an introduction to the writer's article *New division algebras*, in these Transactions, vol. 28 (1926), pp. 207–34 (cited as I), and to current literature based on it. We postpone extensions to non-Abelian equations whose roots have four or more generators.

2. **The algebra $\Gamma$ derived from a given algebra $\Sigma$.** The general theory requires a special case, which serves well as an introductory example.

Note that if $i$ is one root of $x^2-sx+p=0$, the second root is $\theta(i)=s-i$, while $\theta[\theta(i)]=\theta(s-i)=s-(s-i)=i$.

But quadratic equations are only the simplest cases of cyclic equations $f(x)=0$ of degree $p$ which are irreducible in a field $F$ and whose roots are related as follows. If $i$ is one root, it has another root $\theta(i)$ which is a polynomial in $i$ with coefficients in $F$. Using the notation of iteratives (which are not powers), we write $\theta^2(i)$ for $\theta[\theta(i)]$, $\theta^3(i)$ for $\theta[\theta^2(i)]$, etc. Then the roots of the cyclic equation are $i, \theta(i), \theta^2(i), \cdots, \theta^{p-1}(i)$, and we have $\theta^p(i)=i$.

The field $F(i)$ is composed of all rational functions of $i$ with coefficients in $F$, while each such function is known to be equal to a polynomial in $i$ with coefficients in $F$. Hence $F(i)$ may be regarded as an algebra over $F$ having the basis $1, i, i^2, \cdots, i^{p-1}$.

* Presented to the Society, February 22, 1930; received by the editors in December, 1929.

For our introductory example, we take $F(i)$ as algebra $\Sigma$. Each element $A$ of $\Sigma$ is therefore a polynomial $A(i)$ in $i$ with coefficients in $F$. Since $i$ and $\theta$ are roots of the same equation $f(x)=0$, irreducible in $F$, $A(i)B(i)=C(i)$ implies $A(\theta)B(\theta)=C(\theta)$. To each $A=A(i)$, let correspond $A'=A(\theta)$. Hence

$$(1) \qquad (AB)' = A'B', \quad (A+B)' = A'+B' \quad \text{for all } A \text{ and } B \text{ in } \Sigma.$$

We denote $(A')'$ by either $A''$ or $A^{(2)}$. In general, let $A^{(k)}$ denote the element which corresponds to $A^{(k-1)}$. Here

$$A^{(2)} = A[\theta^2(i)], \cdots, A^{(p)} = A[\theta^p(i)] = A(i) = A.$$

Hence for every $\gamma$ in the present commutative algebra $\Sigma$, we have

$$(2) \qquad A^{(p)}\gamma = \gamma A \quad \text{for all } A \text{ in } \Sigma.$$

To simplify the notations, let $p=3$. If $A$, $B$, $C$ are any polynomials in $i$, consider the matrix

$$M(A,B,C) = \begin{pmatrix} A & B & C \\ C'\gamma & A' & B' \\ B''\gamma & C''\gamma & A'' \end{pmatrix}.$$

In case $\gamma=\gamma'$ (whence $\gamma$ is in $F$), we find that the product of any two matrices $M(A, B, C)$ is a third such matrix. Likewise for their sum and for the product of $M$ by any number in $F$. Hence the totality of such matrices $M$ is an associative algebra over $F$. Special cases of $M$ are

$$(A) = \begin{pmatrix} A & 0 & 0 \\ 0 & A' & 0 \\ 0 & 0 & A'' \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \gamma & 0 & 0 \end{pmatrix}, \quad T^2 = \begin{pmatrix} 0 & 0 & 1 \\ \gamma & 0 & 0 \\ 0 & \gamma & 0 \end{pmatrix}.$$

We see that

$$M(A,B,C) = (A) + (B)T + (C)T^2, \quad T(A) = (A')T, \quad T^3 = (\gamma).$$

The algebra whose elements are the matrices $(A)$ is evidently equivalent to the algebra with the elements $A$. It is now a simple step to infer that the algebra whose elements are the matrices $M(A, B, C)$ is equivalent to an algebra $\Gamma$ whose elements are $A+BE+CE^2$, with $E^3=\gamma$, $EA=A'E$.

We readily extend* our proof from 3-rowed to $p$-rowed matrices and obtain

* I, pp. 217–19. Simpler in *Algebren und ihre Zahlentheorie*, p. 59.

THEOREM 1. *Let $\Sigma$ be any given associative algebra of order $p$. To each element $A$ of $\Sigma$ let correspond an element $A'$ also in $\Sigma$ such that* (1) *and* (2) *hold, where $\gamma=\gamma'$ is a self-corresponding element of $\Sigma$. Then there exists an associative algebra $\Gamma$ whose elements are*

$$A_0 + A_1E + \cdots + A_{p-1}E^{p-1}, \tag{3}$$

*where the $A$'s range independently over $\Sigma$, such that*

$$E^p = \gamma, \quad E^rA = A^{(r)}E^r \qquad (r = 1,2, \cdots; \text{any } A \text{ in } \Sigma). \tag{4}$$

By means of (4), the product of any two elements (3) can evidently be expressed in the form (3).

If in our example of a cyclic equation we replace $p$, $\gamma$, $E$ by $q$, $g$, $j$, respectively, we obtain

THEOREM 2. *Let an equation be irreducible in a field $F$ and have the roots $i, \theta(i), \theta^2(i)=\theta[\theta(i)], \cdots, \theta^{q-1}(i)$, where $\theta(i)$ is a polynomial in $i$ with coefficients in $F$ whose $q$th iterative $\theta^q(i)$ is equal to $i$. Then the algebra over $F$ with the basis* $i^mj^n (m, n=0, 1, \cdots, q-1)$, where*

$$j^q = g(i), \; j^rP(i) = P[\theta^r(i)]j^r \qquad (r = 1,2, \cdots) \tag{5}$$

*for every polynomial $P(i)$, is associative if and only if*

$$g(i) = g(\theta). \tag{6}$$

3. **Algebras defined by an equation whose roots have two generators.** The roots of $\xi^4+c\xi^2+1=0$ are $i, -i, 1/i, -1/i$. Write $\theta(i)=-i$, $\phi(i)=1/i = -i^3-ci$. Then the roots are $i, \theta(i), \phi(i)$ and $\theta[\phi(i)]=\phi[\theta(i)]$. As a generalization, let $f(\xi)=0$ be an equation of degree $pq$ which is irreducible in a field $F$ and has the roots

$$\phi^r[\theta^s(i)] \qquad (r = 0,1, \cdots, p-1; s = 0,1, \cdots, q-1), \tag{7}$$

where $r$ and $s$ denote iteratives and not exponents, while $\phi(\xi)$ and $\theta(\xi)$ are polynomials with coefficients in $F$. Further, let

$$\begin{gathered} \theta^q(i) = i, \quad \phi^p(i) = \theta^s(i), \\ \theta[\phi(i)] = \phi[\theta^s(i)], \end{gathered} \tag{8}$$

where $q$ and $p$ are the least positive integers for which relations of type $(8_1)$ and $(8_2)$ hold.

Let $\Sigma$ be the algebra over $F$ having the basis $i^mj^n (m=0, 1, \cdots, pq-1; n=0, 1, \cdots, q-1)$ such that (5) and $f(i)=0$ hold. This $\Sigma$ may be obtained from Theorem 2 by replacing $F$ by the field $F_1$ derived by adjoining to $F$

* Before the associative law is proved $j^3, j^4, \cdots$ are to be replaced by $j_3, j_4, \cdots$.

all the elementary symmetric functions of $i, \theta(i), \cdots, \theta^{q-1}(i)$. Hence the latter are roots of an equation of degree $q$ with coefficients in $F_1$, which is irreducible in $F_1$ by I, §6. Since any polynomial in $i$ over $F_1$ is equal to a polynomial in $i$ over $F$, the algebra over $F_1$ in Theorem 2 may be regarded as our algebra $\Sigma$ over $F$. We assume (6). Then $\Sigma$ is associative.

We shall construct an algebra $\Gamma$ by means of Theorem 1. Each element of $\Sigma$ is of the form

$$A = \sum_{k=0}^{q-1} f_k(i) j^k, \tag{9}$$

where the $f_k(i)$ are polynomials in $i$ of degrees $\leqq pq-1$ with coefficients in $F$. To $A$ we let correspond*

$$A' = \sum_{k=0}^{q-1} f_k(\phi) j'^k, \quad j' = \alpha j^x. \tag{10}$$

In particular, $i' = \phi(i)$. By means of the relations

$$j^q = g(i), \; ji = \theta(i) j, \; f(i) = 0, \tag{11}$$

we can evidently reduce any product $AB$ to $C$, where also $B$ and $C$ are of type (9). Performing the same operations on $i'$ and $j'$ that we did on $i$ and $j$, we conclude that $A'B'$ can be reduced to $C'$ by means of the relations

$$j'^q = g(\phi), \; j'\phi(i) = \theta[\phi(i)] j', \; f(\phi) = 0. \tag{12}$$

In other words, (12) imply the desired property $(1_1)$.

When do relations (12) hold? In $(12_2)$ we insert the value (10) of $j'$, apply (5) for $r = x$, and obtain a relation which follows at once from $(8_3)$. We shall next prove by induction on $m$ that

$$\begin{aligned} j'^m &= \pi_m j^{mx}, \\ \pi_m(i) &= \alpha\alpha(\theta^x)\alpha(\theta^{2x}) \cdots \alpha(\theta^{(m-1)x}). \end{aligned} \tag{13}$$

For, if we multiply the second member of $(13_1)$ on the right by $\alpha j^x$ (or $j'$) and note that $j^{mx}\alpha = \alpha(\theta^{mx}) j^{mx}$ by (5), we get $\pi_{m+1} j^{(m+1)x}$. Hence (13) holds also when $m$ is replaced by $m+1$. The case $m = q$ of (13) shows that $(12_1)$ is equivalent to

$$\pi_q g^x = g(\phi). \tag{14}$$

* We desire that $\Sigma$ be enlarged to $\Gamma$ by an extender $E = k$ such that $kA = A'k$ by (4) and such that (32) shall hold. Hence $i' = \phi(i), j' = \alpha j^x$.

In §2 we require a self-corresponding element $\gamma=\gamma'$ of $\Sigma$. Let* $\gamma=\beta(i)j^e$. By (13),

$$\gamma' = \beta(\phi)j'^e = \beta(\phi)\pi_e j^{ex}.$$

By $(8_3)$ we see by induction that $\theta^k(\phi)=\phi(\theta^{kx})$. Take $k=e$ and note that $(8_2)$ implies $\theta^e(\phi)=\phi^p(\phi)=\phi(\phi^p)=\phi(\theta^e)$. Hence $\theta^{ex}=\theta^e$, $ex=e+c_1q$, where $c_1$ is an integer.† Hence $\gamma'=\gamma$ if and only if

$$\beta(\phi)\pi_e g^{c_1} = \beta, \quad c_1 = (ex-e)/q = \text{integer}. \tag{15}$$

It remains only to satisfy condition (2). If it holds for $A=B$ and for $A=C$, (1) gives

$$\begin{aligned}(BC)^{(p)}\gamma &= B^{(p)}C^{(p)}\gamma \\ &= B^{(p)}\gamma C = \gamma BC,\end{aligned}$$

whence (2) holds also for $A=BC$, and likewise for $A=B+C$. By (9) it therefore suffices to prove (2) for the cases $A=j$ and $A=a(i)$, a polynomial in $i$. For the latter case,

$$a^{(p)}(i) = a(\phi^p) = a(\theta^e)$$

by $(8_2)$. By (5),

$$\begin{aligned}\gamma a &= \beta j^e a \\ &= \beta a(\theta^e)j^e = a^{(p)}(i)\gamma.\end{aligned}$$

Hence there remains only the case $A=j$ of (2), denoted by (2′). By induction on $r$, using (13) with $m=x^r$, we get

$$j^{(r)} = \alpha(\phi^{r-1})\pi_x(\phi^{r-2})\pi_{x^2}(\phi^{r-3})\cdots\pi_{x^{r-1}}j^{x^r}. \tag{16}$$

Take $r=p$. Then (2′) requires‡ that

$$x^p = qc_2 + 1, \quad c_2 \text{ an integer}. \tag{17}$$

Hence (2′) is equivalent to

$$\alpha(\phi^{p-1})\pi_x(\phi^{p-2})\pi_{x^2}(\phi^{p-3})\cdots\pi_{x^{p-2}}(\phi)\pi_{x^{p-1}}g^{c_2}\beta(\theta) = \beta. \tag{18}$$

---

* In this choice we are guided by $(4_1)$ and $(8_2)$, the fact that $j$ and $E$ are associated with the respective roots $\theta$ and $\phi$, and finally by Lemma 2 of I, p. 212.

† This follows also from $\gamma'=\gamma$ and the fact that $q$ is a minimum in $(11_1)$.

‡ Also by (8). To prove $\theta(\phi^s)=\phi^s(\theta^{x^s})$ by induction on $s$, replace $i$ by $\phi$ and apply the earlier result $\theta^k(\phi)=\phi(\theta^{kx})$ for $k=x^s$. Take $s=p$ and note that $\theta(\phi^p)=\theta(\theta^e)=\theta^e(\theta)=\phi^p(\theta)$. Hence $\theta^{x^p}=\theta$.

By Theorem 1 with $E$ replaced by $k$, we now have

THEOREM 3. *Let an equation $f(\xi)=0$ be of degree $pq$, be irreducible in a field $F$, and have the roots* (7) *satisfying relations* (8). *Consider the algebra over $F$ having the basis $i^m j^n k^t (m=0, 1, \cdots, pq-1;\ n=0, 1, \cdots, q-1;\ t=0, 1, \cdots, p-1)$ with $f(i)=0$,* (5), *and**

$$k^p = \beta(i)j^s, \quad k^r P(i) = P(\phi^r)k^r, \quad k^r j^s = [j^{(r)}]^s k^r, \tag{19}$$

*where $j^{(r)}$ has the value* (16). *This algebra is associative if and only if conditions*† (6), (14), (15), *and* (18) *hold.*

Simplifications and interpretations of these conditions will be given in §§7, 8. Theorem 3 is equivalent to that proved by J. Williamson in these Transactions, vol. 30 (1928), pp. 112–5, by the method used by the writer for the case $x=q-1$ in I, pp. 228–34. That method is much more complicated than the present one since it examined the conditions for (1) for all $A$ and $B$, whereas it suffices to treat the cases (12).

4. **Algebras defined by Abelian equations.** The cyclic equation in §2 and the quartic in §3 are examples of Abelian equations. The general one is defined in

THEOREM 4. *Let an Abelian equation $f(x)=0$ be of degree $d=p_1 p_2 \cdots p_n$, irreducible in a field $F$, and have the roots*

$$\psi_1^{e_1}\{\psi_2^{e_2}[\cdots \psi_n^{e_n}(i)]\} \qquad \begin{pmatrix} e_1 = 0, 1, \cdots, p_1 - 1 \\ \cdots\cdots\cdots\cdots \\ e_n = 0, 1, \cdots, p_n - 1 \end{pmatrix},$$

*where each $\psi_j(i)$ is a polynomial in $i$ with coefficients in $F$, $\psi^e(t)$ denotes the $e$th iterative of $\psi(t)$, such that the $p_r$th iterative of $\psi_r(i)$ is equal to $i$, and*

$$\psi_j[\psi_k(i)] = \psi_k[\psi_j(i)] \qquad (j, k = 1, \cdots, n).$$

*Consider the algebra $A_n$ over $F$ with the basis*

$$i^h j_1^{e_1} \cdots j_n^{e_n} \qquad (h = 0, 1, \cdots, d-1;\ e_s = 0, 1, \cdots, p_s - 1),$$

*such that, for every polynomial $P(i)$,*

$$j_r^{p_r} = g_r(i), \quad j_r^m P(i) = P[\psi_r^m(i)]\, j_r^m \qquad (r = 1, \cdots, n;\ m = 1, 2, \cdots), \tag{20}$$

$$j_s^k j_r^m = T_m(i) T_m(\psi_s) \cdots T_m(\psi_s^{k-1}) j_r^m j_s^k \qquad (s, r = 1, \cdots, n;\ s > r), \tag{21}$$

* After the theorem is proved, we may replace (5) and (19) by their special cases (11) and (32), which with the associative law imply the former.

† We do not include the facts that $c_1$ and $c_2$ in (15) and (17) are integers, since they follow from the hypotheses (8).

$$(22) \qquad T_m(i) = T_{m,s,r} = \alpha(i)\alpha(\psi_r)\alpha(\psi_r^2)\cdots\alpha(\psi_r^{m-1}), \quad \alpha \equiv \alpha_{s,r}.$$

*Then $A_n$ is associative if and only if*

$$\alpha_{t,s}\alpha_{s,r}\alpha_{t,r}(\psi_s) = \alpha_{t,s}(\psi_r)\alpha_{s,r}(\psi_t)\alpha_{t,r} \qquad (t > s > r),$$

$$T_{p_r,t,r}g_r(i) = g_r(\psi_t), \quad \alpha_{t,r}\alpha_{t,r}(\psi_t)\cdots\alpha_{t,r}(\psi_t^{p_t-1})g_t(\psi_r) = g_t \qquad (t > r),$$

*and $g_t(i) = g_t(\psi_t)$, for $r, s, t = 1, \cdots, n$.*

After the theorem is proved, we may replace $(20_2)$ and (21) by the special cases

$$j_r i = \psi_r(i)j_r, \quad j_s j_r = \alpha_{s,r}(i)j_r j_s$$

of them, since these and the associative law imply the former.

We shall prove Theorem 4 by induction on $n$, assuming* that $\Sigma = A_{n-1}$ is associative. Let $R$ denote the relations (20) and (21) with subscripts restricted to $1, \cdots, n-1$. Let $R'$ denote the relations derived from $R$ by replacing $i$ and $j_r$ by

$$(23) \qquad i' = \psi_n(i), \quad j_r' = \alpha_{n,r}(i)j_r.$$

By means of relations $R$ we can evidently express any element of $\Sigma$ in the form

$$(24) \qquad A = \sum_{e_1=0}^{p_1-1} \cdots \sum_{e_{n-1}=0}^{p_{n-1}-1} a_{e_1,\cdots,e_{n-1}}(i)j_1^{e_1}\cdots j_{n-1}^{e_{n-1}}.$$

Similarly, by means of $R$, we can reduce any product $AB$ to $C$, where $B$ and $C$ are of type (24) with $a$'s replaced by $b$'s and $c$'s, respectively. Let $A'$ denote the sum (24) with $i$ and $j_r$ replaced by (23). Hence $A'B'$ can be reduced to $C'$ by means of relations $R'$ (since the operations on the accented letters are the same as those used on unaccented letters). In other words, $R'$ imply property (1).

By the remark below the theorem, $R'$ follow from the special cases

$$(25) \qquad j_r'^{p_r} = g_r(\psi_n), \quad j_r'\psi_n = \psi_r(\psi_n)j_r', \quad j_s'j_r' = \alpha_{s,r}(\psi_n)j_r'j_s' \qquad (s, r = 1, \cdots, n-1; s > r)$$

of them, in which the values of the $j'$ in (23) are to be inserted. From $(25_2)$ and $(20_2)$ with $m = 1$, we get

$$\alpha_{n,r}\psi_n[\psi_r(i)]j_r = \psi_r[\psi_n(i)]\alpha_{n,r}j_r$$

which is true by the assumption in the theorem. Similarly, $(25_3)$ becomes

* And employing the analog of the paragraph following (8). Compare the second paragraph of §5.

$$\alpha_{n,s}\alpha_{n,r}(\psi_s)j_sj_r = \alpha_{s,r}(\psi_n)\alpha_{n,r}\alpha_{n,s}(\psi_r)j_rj_s.$$

By the relation below the theorem, this holds if and only if

$$\alpha_{n,s}\alpha_{n,r}(\psi_s)\alpha_{s,r} = \alpha_{s,r}(\psi_n)\alpha_{n,r}\alpha_{n,s}(\psi_r) \qquad (r,s = 1, \cdots, n-1; s > r). \tag{26}$$

By induction on $m$

$$j_r'^m = T_{m,n,r}j_r^m.$$

Hence $(25_1)$ is equivalent to

$$T_{p_r,n,r}g_r(i) = g_r(\psi_n). \tag{27}$$

Since the desired formula $(4_1)$ is here $(20_1)$ with $r=n$, we must choose $\gamma=g_n(i)$. Then $\gamma=\gamma'$ becomes

$$g_n(i) = g_n(\psi_n). \tag{28}$$

Also, $p=p_n$ in (2), which requires proof only when $\gamma=g_n\neq 0$. If $A$ is a polynomial $a(i)$ in $i$, (2) holds since

$$\psi_n^{p_n}(i) = i, \quad g_nag_n^{-1} = a(i) = a(\psi_n^{p_n}) = a^{(p_n)}.$$

As below (15), it remains only to consider (2) for $A=j_r$, $r<n$. By (20), $j_rg_n=g_n(\psi_r)j_r$. Hence shall

$$g_nj_rg_n^{-1} = \frac{g_n}{g_n(\psi_r)}j_r = j_r^{(p)}$$

for $p=p_n$. By induction on $p$,

$$j_r^{(p)} = \alpha\alpha(\psi_n)\cdots\alpha(\psi_n^{p-1})j_r, \quad \alpha \equiv \alpha_{n,r}.$$

Hence (2) holds for $A=j_r$ if and only if

$$\alpha(i)\alpha(\psi_n)\cdots\alpha(\psi_n^{p_n-1}) = \frac{g_n}{g_n(\psi_r)}, \quad \alpha \equiv \alpha_{n,r} \quad (r = 1, \cdots, n-1). \tag{29}$$

Under the assumption that $A_{n-1}$ is associative, we have now proved that $A_n$ is associative if and only if conditions (26)-(29) hold. But $A_1$ is of the type in Theorem 2 and hence is associative if and only if $g_1(i)=g_1(\psi_1)$. This completes the proof of Theorem 4. It was first proved (in manuscript) two years ago by J. S. Georges after very long reductions of all conditions arising from (1).

5. **Algebras defined by an equation whose roots have three generators.** Let $f(\xi)=0$ be an equation of degree $pqs$ which is irreducible in $F$ and has the roots

$$\psi^m\{\phi^n[\theta^r(i)]\} \qquad (m < s, n < p, r < q). \tag{30}$$

Let the roots satisfy relations (8) and

$$\psi^s = \phi^b(\theta^a),\ \theta(\psi) = \psi[\phi^v(\theta^v)],\ \phi(\psi) = \psi[\phi^z(\theta^w)], \tag{31}$$

from which the argument $i$ has been suppressed, while $s$ is a minimum.

Let $\Sigma$ be the algebra over $F$ having the basis $i^m j^n k^t$ ($m \leqq pqs-1$, $n \leqq q-1$, $t \leqq p-1$) such that (5), (19), and $f(i)=0$ hold. As in the paragraph below (8), $\Sigma$ may be obtained from Theorem 3 by replacing $F$ by the field derived by adjoining to $F$ all elementary symmetric functions of the roots (7). We assume (6), (14), (15), and (18). Then $\Sigma$ is associative. We see that (5) and (19) are consequences of the associative law and the set $R$ of relations given by (11) and the following cases of (19):

$$k^p = \beta(i)j^e,\ ki = \phi(i)k,\ kj = \alpha(i)j^z k. \tag{32}$$

Let $R'$ denote the set of like relations (11′) and (32′) with $i, j, k$ replaced by $i', j', k'$, which are defined, in accord with (4) and (31), to be

$$i' = \psi(i),\ J \equiv j' = \epsilon(i)j^v k^v,\ K \equiv k' = \delta(i)j^w k^z. \tag{33}$$

By the usual argument, relations $R'$ imply property (1). We insert the values (33) into the five relations $R'$. By ($5_2$) and ($19_2$), we see that ($11_2'$) becomes

$$\epsilon\psi[\phi^v(\theta^v)]j^v k^v = \epsilon\theta(\psi)j^v k^v,$$

which is true by ($31_2$). Similarly, ($32_2'$) follows from ($31_3$).

Using $\pi_m$ in (13), we see by induction on $n$ that

$$kj^n = \pi_n(i)j^{z^n}k. \tag{34}$$

By induction on $m$,

$$k^m j^n = B_{m,n}j^{z^m n}k^m, \tag{35}$$

$$B_{m,n}(i) = \pi_n(\phi^{m-1})\pi_{z^n}(\phi^{m-2})\pi_{z^{2n}}(\phi^{m-3})\cdots\pi_{z^{m-1}n}(i). \tag{36}$$

By ($5_2$) and ($19_2$),

$$KJ = \delta\epsilon[\phi^z(\theta^w)]j^w k^z j^v k^v.$$

Applying (35) with $m=z$, $n=y$, and then ($5_2$), we get

$$KJ = \delta\epsilon[\phi^z(\theta^w)]B_{z,y}(\theta^w)j^{w+z^y v}k^{z+v}. \tag{37}$$

We employ the abbreviations

$$h(t) = y + x^v y + x^{2v}y + \cdots + x^{(t-1)v}y,\ h(0) = 0, \tag{38}$$

$$C_d(i) = \epsilon\epsilon(\phi^v\theta^v)\epsilon(\phi^{2v}\theta^{h(2)})\epsilon(\phi^{3v}\theta^{h(3)})\cdots\epsilon(\phi^{(d-1)v}\theta^{h(d-1)}) \tag{39}$$
$$\times B_{v,y}(\theta^v)B_{2v,y}(\theta^{h(2)})B_{3v,y}(\theta^{h(3)})\cdots B_{(d-1)v,y}(\theta^{h(d-1)}),$$

whence $C_1=\epsilon$. Hence if $d=1$,

$$J^d = C_d j^{h(d)} k^{dv}. \tag{40}$$

Assuming (40), we obtain the like result with $d$ replaced by $d+1$ by using (37) with $\delta$, $w$, $z$ replaced by $C_d$, $h(d)$, $dv$, respectively, whence $K$ is replaced by $J^d$. This proves (40) by induction.

The case $d=q$ of (40) shows that $(11_1')$ is equivalent to

$$C_q j^{h(q)} k^{qv} = g(\psi). \tag{41}$$

From $(19_1)$ by induction on $n$, we get

$$k^{np} = \beta_n j^{en}, \ \beta_n(i) = \beta\beta(\theta^e)\cdots\beta(\theta^{e(n-1)}). \tag{42}$$

Since no lower than the $p$th power of $k$ is a polynomial in $i$ and $j$, (41) requires that

$$qv = c_3 p, \quad c_3 \text{ an integer}. \tag{43}$$

Replacing $k^{qv}$ by its value (42), and using (5), we see that the new form of (41) involves $j$ with the exponent

$$h(q) + ec_3 = c_4 q, \ c_4 \text{ an integer}, \tag{44}$$

and hence that (41) is equivalent to

$$C_q \beta_{c_4}(\theta^{h(q)}) g^{c_4} = g(\psi). \tag{45}$$

By a change of notation in (37) we see by (40) that

$$J^z K = C_z \delta(\phi^{xv}\theta^{h(x)}) B_{xv,w}(\theta^{h(x)}) j^{\xi} k^{xv+z},$$

$$\xi = h(x) + wx^{xv}. \tag{46}$$

Thus $\xi$ is $t(x, 1)$ in the notation (58). By $(32_3')$, $\alpha(\psi)J^z K = KJ$. By the exponents of $k$, we have

$$xv - v = c_5 p. \tag{47}$$

We apply (42) with $n$ replaced by $c_5$. By the resulting terms in $j$,

$$\xi + ec_5 - w - x^z y = c_6 q. \tag{48}$$

The resulting condition is

$$\alpha(\psi) C_z \delta(\phi^{xv}\theta^{h(x)}) B_{xv,w}(\theta^{h(x)}) \beta_{c_4}(\theta^{\xi}) g^{c_4} = \delta\epsilon(\phi^z\theta^w) B_{z,y}(\theta^w). \tag{49}$$

We obtain $K^d$ from $J^d$ by the replacement in (51). Let

$$H(t) = w + x^z w + x^{2z} w + \cdots + x^{(t-1)z} w, \quad H(0) = 0, \tag{50}$$

(51) $D_d(i)$ be derived from (39) by replacing $\epsilon$, $y$, $v$, $h$ by $\delta$, $w$, $z$, $H$.

In these notations, we have

$$K^d = D_d j^{H(d)} k^{dz}. \tag{52}$$

Hence $(32_1')$ becomes

$$D_p j^{H(p)} k^{pz} = \beta(\psi) C_e j^{h(e)} k^{ev}.$$

By the terms in $k$ and those in $j$ after reductions by (42), we must have

$$ev = c_7 p, \quad H(p) + ez - h(e) - ec_7 = c_8 q, \tag{53}$$

$$D_p \beta_z(\theta^{H(p)}) g^{c_8} = \beta(\psi) C_e \beta_{c_7}(\theta^{h(e)}). \tag{54}$$

We have now examined all the relations (11′) and (32′).

A comparison of $(4_1)$ with $(31_1)$ leads us to choose

$$\gamma = \sigma(i) j^a k^b, \tag{55}$$

and later to identify the $p$ in $(4_1)$ with our $s$. By (40), (52), and (37) in changed notations, we get

$$J^m K^n = F_{m,n} j^{t(m,n)} k^{mv+nz}, \tag{56}$$

$$F_{m,n}(i) = C_m D_n [\phi^{mv}(\theta^{h(m)})] B_{mv, H(n)}(\theta^{h(m)}), \tag{57}$$

$$t(m,n) = h(m) + x^{mv} H(n). \tag{58}$$

We take $m=a$, $n=b$. Comparing the exponents of $k$ in $\gamma'=\gamma$, we get

$$av + bz - b = c_9 p. \tag{59}$$

By (42),

$$k^{av+bz} = \beta_{c_9} j^{ec_9} k^b.$$

Comparing the exponents of $j$ in $\gamma'=\gamma$, we get

$$t(a,b) + ec_9 - a = c_{10} q. \tag{60}$$

Hence the conditions for $\gamma'=\gamma$ are (59), (60), and

$$\sigma(\psi) F_{a,b} \beta_{c_9}(\theta^{t(a,b)}) g^{c_{10}} = \sigma(i). \tag{61}$$

Finally, we consider (2), viz.,

$$A^{(s)}\gamma = \gamma A \text{ for all } A \text{ in } \Sigma. \tag{62}$$

This holds if $A$ is any polynomial $P(i)$ in $i$. For, by $(19_2)$, $(5_2)$, and $(31_1)$,

$$\gamma P = \sigma j^a P(\phi^b) k^b = \sigma P[\phi^b(\theta^a)] j^a k^b = \sigma P(\psi^s) j^a k^b = P^{(s)}\gamma.$$

By the remarks below (15), it remains only to find the conditions that (62) shall hold for the values $j$ and $k$ of $A$. By (33),

$$j'' = \epsilon(\psi)J^{y}K^{v}, \qquad k'' = \delta(\psi)J^{w}K^{z}. \tag{63}$$

Hence for two sets of values of $m$ and $n$ we need a formula of type

$$(J^{m}K^{n})^{(r)} = G_{m,n,r}(i)j^{f(m,n,r)}k^{l(m,n,r)}. \tag{64}$$

Accenting each member and applying (33) and (56), we get a relation which is an identity in $j$ and $k$ if

$$l(m,n,r+1) = vf + zl, \quad f(m,n,r+1) = t(f,l), \tag{65}$$

$$G_{m,n,r+1} = G_{m,n,r}(\psi)F_{f,l}, \tag{66}$$

in which the arguments of $f$ and $l$ in the second members are $m$, $n$, $r$. Since we may identify (64) for $r=0$ with (56),

$$G_{m,n,0} = F_{m,n}, \quad f(m,n,0) = t(m,n), \quad l(m,n,0) = mv + nz. \tag{67}$$

For these initial values, (65) and (66) serve as recursion formulas to determine the $f$, $l$, $G$ uniquely.

By $(63_1)$ and (64) with $r=s-2$ we get $j^{(s)}$. Its product on the right by $\gamma$ in (55) is found by the rule (37). Hence for $V$ in (70),

$$j^{(s)}\gamma = Vj^{f+x^{l}a}k^{l+b}, \quad \gamma j = \sigma B_{b,1}(\theta^{a})j^{a+x^{b}}k^{b},$$

the second from (35). These shall be equal by (62) with $A=j$. Comparing exponents of $k$ and, after using (42), exponents of $j$, we get

$$l = c_{11}p, \quad f + x^{l}a + ec_{11} - a - x^{b} = c_{12}q, \tag{68}$$

$$V\beta_{c_{11}}(\theta^{f+x^{l}a})g^{c_{12}} = \sigma B_{b,1}(\theta^{a}), \tag{69}$$

$$V = \epsilon(\psi^{s-1})G\sigma(\phi^{l}\theta^{f})B_{l,a}(\theta^{f}), \tag{70}$$

where the arguments of $f$ and $l$ and the subscripts of $G$ are $y$, $v$, $s-2$.

Similarly, starting with $(63_2)$, we find that (62) holds with $A=k$ if and only if

$$l - 1 = c_{13}p, \quad f + x^{l}a + ec_{13} - a = c_{14}q, \tag{71}$$

$$W\beta_{c_{13}}(\theta^{f+x^{l}a})g^{c_{14}} = \sigma, \tag{72}$$

where $W$ is derived from (70) by replacing $\epsilon$ by $\delta$, and the arguments of $f$ and $l$ and the subscripts of $G$ are now $w$, $z$, $s-2$.

By Theorem 1 with $p$ replaced by $s$, we now have

THEOREM 5. *Let an equation $f(\xi)=0$ be of degree $pqs$, be irreducible in a field $F$, and have the roots* (30) *satisfying relations* (8) *and* (31). *Consider the algebra over $F$ having the basis $i^m j^n k^t E^c (m=0,\cdots,pqs-1;\ n=0,\cdots,q-1;\ t=0,\cdots,p-1;\ c=0,\cdots,s-1)$ with $f(i)=0$,* (5), (19), *and**

$$
\begin{aligned}
&E^s = \sigma(i) j^a k^b, && E^r P(i) = P(\psi^r) E^r,\\
&E^r j^s = [j^{(r)}]^s E^r, && E^r k^s = [k^{(r)}]^s E^r,
\end{aligned}
\tag{73}
$$

*where*

$$
j^{(r)} = \epsilon(\psi^{r-1})(J^u K^v)^{(r-2)}, \quad k^{(r)} = \delta(\psi^{r-1})(J^w K^z)^{(r-2)}
$$

*are found by* (64), *while their sth powers may be found by* (40) *with altered parameters. This algebra is associative if and only if conditions* (6), (14), (15), (18), (45), (49), (54), (61), (69), *and* (72) *hold.*

We do not include as conditions the facts that $c_1,\cdots,c_{14}$ are integers in (15), (17), (43), (44), (47), (48), (53), (59), (60), (68), and (71), since we shall prove in §6 that they follow from the hypotheses (8) and (31).

6. **Groups.** Let $r_1(i),\cdots,r_d(i)$ denote the $d$ roots† (30) arranged in an arbitrarily chosen order. Let $j$ and $k$ be any of the numbers $1,\cdots,d$. By means of (8) and (31), we can evidently reduce the function $r_k[r_j(i)]$ to the form (30) and hence obtain a certain root $r_l(i)$. If $j$ is fixed, but $k$ takes the values $1,\cdots,d$, then $l$ takes the same values rearranged. Hence we obtain a substitution $S_j$ on the $d$ roots which is said to replace $r_k$ by $r_l$.

Similarly, if $t$ is fixed and $r_l[r_t(i)]=r_f$, we obtain a substitution $S_t$ which replaces $r_l$ by $r_f$. Let $r_j[r_t(i)]=r_c$. Then

$$
r_k(r_c) = r_k[r_j(r_t)] = r_l(r_t) = r_f,
$$

which defines a substitution $S_c$ replacing $r_k$ by $r_f$. It is called the product of $S_j$ by $S_t$ and denoted by $S_jS_t$.

We therefore have a group of $d$ substitutions $S$ such that

$$
S_jS_t = S_c \quad \text{if} \quad r_j[r_t(i)] = r_c(i),
\tag{75}
$$

while $S_j$ is uniquely determined by the root $r_j(i)$.

We return to the notation (30) for the roots. Let $\Theta$, $\Phi$, $\Psi$ denote the

---

* After the theorem is proved, we may replace (5), (19), and the last three of (73) by their special cases (11), (32) and

$$
Ei=\psi(i)E, \quad Ej=j'E, \quad Ek=k'E, \quad j' \text{ and } k' \text{ in (33)}.
\tag{74}
$$

In fact, the latter and the associative law imply the former.

† The same discussion applies to (7), to the Abelian equation in Theorem 4, and to any equation whose roots are all rational functions of one root.

substitutions determined by the roots $\theta(i)$, $\phi(i)$, $\psi(i)$, respectively. By (8), (31), and (75), we get*

$$\begin{aligned} &\Theta^q = 1, \quad \Phi^p = \Theta^e, \quad \Phi^{-1}\Theta\Phi = \Theta^z, \\ &\Psi^s = \Phi^b\Theta^a, \quad \Psi^{-1}\Theta\Psi = \Phi^v\Theta^v, \quad \Psi^{-1}\Phi\Psi = \Phi^z\Theta^w, \end{aligned} \tag{76}$$

in which $q$, $p$, $s$ cannot be replaced by smaller positive integers.

Write $j$, $k$, $E$ for the inverses of $\Theta$, $\Phi$, $\Psi$, respectively, and define $J$, $K$. Taking the inverse of each equation (76), we get

$$\begin{aligned} &j^q = 1, \qquad k^p = j^e, \qquad kj = j^z k, \qquad E^s = j^a k^b, \\ &Ej = JE, \quad Ek = KE, \quad J = j^v k^v, \quad K = j^w k^z. \end{aligned} \tag{77}$$

We evidently have the same relation between the parameters $q, p, e, x, \cdots$ for (77) as for (76).

The developments in §5 were based on (11), (32), and (73) as simplified in (74). If we suppress all factors which are polynomials in $i$, we obtain (77) and conclude that all resulting relations in §5 become consequences of (77). The fact that $c_3, \cdots, c_{14}$ in (43), $\cdots$, (71) are all integers is therefore a consequence of (77) and hence of (76) and therefore of (8) and (31).

THEOREM 6. *That $c_1, \cdots, c_{14}$ in* (15), $\cdots$, (71) *are integers are conditions for the existence of the group.*

While the proof applies also to $c_1$ and $c_2$ in (15) and (17), our former verification was more elementary.

7. **Simplifications.** Our conditions involve the product $\pi_m$ in (13). When $m$ is large, the number $m$ of factors may be reduced. Since $\theta^q(i) = i$, evidently $\pi_{n+q} = \pi_n\pi_q$. By induction on $y$,

$$\pi_{n+yq} = \pi_n\pi_q{}^y. \tag{78}$$

Here we may replace $\pi_q$ by its simple value from (14). Hence we may materially reduce the number of factors in condition (18). To this end we write

$$x^2 = x_2 + qy_2, \quad xx_2 = x_3 + qy_3, \quad \cdots, \quad xx_{p-1} = x_p + qy_p \quad (0 \leqq x_k < q). \tag{79}$$

Evidently $x_k \equiv x^k \pmod q$. Thus $x_p = 1$ by (17). Write

$$\alpha_{x_k} = \pi_{x_k}\delta^{y_{k+1}} \tag{80}$$

* We do not use the illuminating fact that if we start with an equation irreducible in $F$ whose roots are rational functions of one root $i$ with coefficients in $F$, and whose Galois group for $F$ is generated by $\Theta$, $\Phi$, $\Psi$ subject to (76), then the roots are (30) and satisfy relations (8) and (31). Similarly for the first line of (76) and the roots (7) satisfying (8).

also when $k=1$, taking $x_1=x$. Then (18) is equivalent to*

$$\alpha(\phi^{p-1})\alpha_{x_1}(\phi^{p-2})\alpha_{x_2}(\phi^{p-3})\cdots\alpha_{x_{p-2}}(\phi)\alpha_{x_{p-1}}=\beta\beta^{-1}(\theta). \tag{81}$$

In §5 we may again simplify $\pi_n$ in (34) by use of (78) and (14) and reduce the exponent $xn$ of $j$ below $q$. The new (35) is automatically simplified. After writing this paper, the author suggested to Miss Dora McFarland that she simplify the further formulas in §5. To do this she reduced at every stage all exponents of $j$ and $k$ below $q$ and $p$, respectively (see the preceding footnote). But the work became far more complicated and the formulas exceedingly complex.

8. **Example, interpretations.** Let $x=q-1$ in §3. We exclude the case $q=2$, since the equation is then Abelian and falls under §4. Then $p$ is even by (17). By the last part of (15), $2e\equiv 0 \pmod q$. But $0\leqq e<q$. Hence

$$\text{either } e=0, \text{ or } q \text{ is even and } e=\tfrac{1}{2}q. \tag{82}$$

Since $x, 2x, \cdots, (q-1)x$ are congruent to $q-1, q-2, \cdots, 1$, modulo $q$,

$$\pi_q=\alpha\alpha(\theta)\alpha(\theta^2)\cdots\alpha(\theta^{q-1}). \tag{83}$$

From (79), (80), and $x^2=1+q(q-2)$, we see that if $k$ is even,

$$x_k=1,\quad y_k=q-2,\quad \alpha_{x_k}=\alpha;$$

if $k$ is odd,

$$x_k=x,\quad y_k=0,\qquad \alpha_{x_k}=\pi_{q-1}g^{q-2}\equiv\chi,$$

$$\chi(i)=\alpha\cdot\alpha(\theta^2)\alpha(\theta^3)\cdots\alpha(\theta^{q-1})g^{q-2}. \tag{84}$$

Hence (81) becomes

$$\alpha(\phi^{p-1})\chi(\phi^{p-2})\alpha(\phi^{p-3})\chi(\phi^{p-4})\cdots\alpha(\phi)\chi(i)=\beta\beta^{-1}(\theta). \tag{85}$$

Now (15) becomes

$$\beta(\phi)=\beta \text{ if } e=0;\ \beta(\phi)\pi_e g^{e-1}=\beta \text{ if } e=\tfrac{1}{2}q. \tag{86}$$

THEOREM 7. *When $x=q-1$ in Theorem 3, the conditions that the algebra be associative are* (6), (14), (85), *and* (86).

A long proof of Theorem 7 was given in I, pp. 228–34.

We shall now simplify (85) and interpret it. Multiply (84) by $\alpha(\theta)g$ and apply (14) and (83); we get

$$\alpha(\theta)g\chi=g(\phi). \tag{87}$$

* We may prove (81) independently of (18). By induction on $r$, $j^{(r)}$ is the product of the left member of (81) with $p$ changed to $r$ by $j^{p-r}$. The case $r=p$ shows that (2) with $A=j$ becomes (81).

In (85) we insert the resulting fractional expression for $\chi$, clear of denominators, and obtain $M\beta(\theta)=L\beta$, where

$$M = \alpha(\phi^{p-1})\alpha(\phi^{p-3})\cdots\alpha(\phi)g(\phi^{p-1})g(\phi^{p-3})\cdots g(\phi),$$
$$L = \alpha(\theta\phi^{p-2})\alpha(\theta\phi^{p-4})\cdots\alpha(\theta)g(\phi^{p-2})g(\phi^{p-4})\cdots g,$$

in which consecutive superior letters differ by 2.

Since $\Phi$ transforms $\Theta$ into its inverse, $\Phi^s$ is commutative with $\Theta$ if $s$ is even. Then $\phi^s$ is commutative with $\theta$ and $\theta^{-1}$. If in $L$ we replace $i$ by $\theta^{-1}\phi$, and apply (6), we get $M$. Hence we may replace (85) by

$$L(i)\beta = L(\theta^{-1}\phi)\beta(\theta). \tag{88}$$

In $(8_3)$ replace $i$ by $\theta$. Hence $\theta[\phi(\theta)]=\phi$, $\phi(\theta)=\theta^{-1}(\phi)$.

Let $e=0$ and write $B(i)$ for $L(i)\beta$. In $(86_1)$ replace $i$ by $\theta$; we get $\beta[\theta^{-1}(\phi)]=\beta(\theta)$. Then (88) gives

$$B(\theta^{-1}\phi) = B(i) \text{ or } B(\theta^{-1}) = B(\phi^{-1}). \tag{89}$$

9. **Normalization of groups with two generators.** Let the generators $\Theta$ and $\Phi$ satisfy the relations in the first line of (76). We may write $e=\epsilon f$, where $f$ is prime to $q$, and all prime factors of $\epsilon$ divide $q$. Some of those primes are denoted by $r_i$ and the others by $s_i$, as follows. We may write $\epsilon=\sigma R$, where $\sigma$ is a product of powers of the $s_i$, and $R$ is a product of powers of the $r_i$, such that the exponent of the highest power of $r_i$ [or $s_i$] which divides $R$ [or $\sigma$] is $>$ [or $\leqq$] the exponent of its power which divides $q$. We may write $q=\rho Q$, where $\rho$ is a product of powers of the $r_i$, no one of which divides $Q$. Hence $\rho$ is a divisor of $R$, and $R/\rho$ is divisible by each $r_i$. Also, each $s_i$ divides $Q$, and $\sigma$ is a divisor of $Q$. Write $E=Q+R/\rho$. Since each $r_i$ divides $R/\rho$, but not $Q$, no $r_i$ divides $E$. If a prime divides $Q$, it is distinct from the $r_i$ and hence is not a factor of $R/\rho$ and therefore not of $E$. Hence $q=\rho Q$ is prime to $E$. Since $E$ and $f$ are both prime to $q$, $\Omega=\Theta^{Ef}$ generates all powers of $\Theta$. Now $E\rho=Q\rho+R\equiv R \pmod{\rho Q=q}$. Multiplication by $\sigma f$ shows that

$$\Omega^{\rho\sigma} = \Theta^{\rho\sigma Ef} = \Theta^{\sigma fR} = \Theta^{e}.$$

By $(76_2)$, $\Phi^p=\Omega^{\rho\sigma}$, and $\rho\sigma$ divides $\rho Q=q$.

THEOREM 8. *We may employ a power of $\Theta$ as a new generator in place of $\Theta$ such that the new $e$ in* (76) *is a divisor of $q$.*

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

# BOUNDARY VALUES OF AN ANALYTIC FUNCTION AND THE TCHEBYCHEFF METHOD OF APPROXIMATION*

BY
J. L. WALSH

**1. Introduction.** If the functions $f(z)$, $f_n(z)$, $n=1, 2, \cdots$, are analytic on and within the unit circle $C$ of the $z$-plane, the relation

$$\lim_{n\to\infty} [\max | f(z) - f_n(z) | , z \text{ on } C] = 0 \tag{1.1}$$

implies the relation

$$\lim_{n\to\infty} f_n(z) = f(z)$$

uniformly for $z$ on and within $C$.

There are, however, cases of approximation formally similar to this—which consist, namely, of approximation on the boundary $C$ of a region $B$ by functions analytic on $B+C$, to a function defined and continuous on $C$—where approximation with an arbitrarily small error is not possible. For instance, the function $f(z)$ to be approximated may be the boundary values of an analytic function *with singularities in* $B$, or the approximating functions $f_n(z)$ may all be required to take on prescribed values different from those of $f(z)$ at definite points interior to $B$; in neither of these cases is (1.1) possible. Natural questions to be raised in these new cases are: (1) What is the minimum value $M$ (or greatest lower bound) of

$$\max | f(z) - f_n(z) | , z \text{ on } C,$$

for all functions $f_n(z)$ analytic interior to $C$ and taking on the prescribed values if any? (2) Does a function $f_0(z)$, taking on the prescribed values if any, exist such that

$$\max | f(z) - f_0(z) | = M, \text{ for } z \text{ on } C?$$

(3) If we have

$$\lim_{n\to\infty} [\max | f(z) - f_n(z) | , z \text{ on } C] = M,$$

where the functions $f_n(z)$ are analytic interior to $C$ and take on the prescribed

* Presented to the Society, March 30, 1929; received by the editors December 16, 1929, and April 7, 1930. An abstract of this paper was published in the Proceedings of the National Academy of Sciences, vol. 15 (1929), pp. 799–802.

values if any, what can be said of the convergence of the sequence $\{f_n(z)\}$? (4) In particular, if we consider as a special sequence the Tchebycheff polynomials for approximation to $f(z)$ on $C$, what can be said of the convergence?

There are two theorems of primary interest which in the simplest cases answer these questions:

THEOREM A. *Let $B$ be an arbitrary simply connected region with the boundary $C$, and let the function $f(z)$ be analytic on $C$, meromorphic on $B+C$. Then there exists a function $f_0(z)$ analytic interior to $B$ such that**

$$\overline{\operatorname{bound}}_{\substack{z\to C\\ z \text{ in } B}} |f(z) - f_0(z)| = M$$

*is less than*

$$\overline{\operatorname{bound}}_{\substack{z\to C\\ z \text{ in } B}} |f(z) - f_n(z)|$$

*or any other function $f_n(z)$ analytic interior to $B$. The expression $|f(z)-f_0(z)|$ approaches the same limit $M$ no matter how $z$ in $B$ approaches $C$. If the functions $f_n(z)$, $n=1, 2, \cdots$, are analytic interior to $B$, then*

$$\lim_{n\to\infty} \left[\overline{\operatorname{bound}}_{\substack{z\to C\\ z \text{ in } B}} |f(z) - f_n(z)|\right] = M$$

*implies*

$$\lim_{n\to\infty} f_n(z) = f_0(z)$$

*for $z$ in $B$, uniformly for $z$ on an arbitrary closed point set interior to $B$.*

*If the region $B$ is limited and if its boundary $C$ is also the boundary of an infinite region, the Tchebycheff polynomial $\omega_n(z)$ of degree $n$ for approximation to $f(z)$ on $C$ is defined as the polynomial of degree $n$ such that*

$$\max |f(z) - \omega_n(z)|, \quad z \text{ on } C,$$

*is less than the corresponding expression formed for any other polynomial of degree $n$; the Tchebycheff polynomial $\omega_n(z)$ exists and is unique. It is then true if $f(z)$ has no singularities other than in $B$ and in the infinite region of which $C$ is the boundary, that*

$$\lim_{n\to\infty} [\max |f(z) - \omega_n(z)|, \quad z \text{ on } C] = M,$$

*from which it follows that*

$$\lim_{n\to\infty} \omega_n(z) = f_0(z) \tag{1.2}$$

*for $z$ in $B$, uniformly on any closed point set interior to $B$.*

* The notation here is intended to imply that we allow $z$ in $B$ to approach $C$ in any way whatever and take the least upper bound of all corresponding limits of $|f(z)-f_0(z)|$.

*If $B$ is the interior of a limited analytic Jordan curve $C$, we denote by $C_R$ the curve which is the locus $|\phi(z)| = R > 1$, where $w = \phi(z)$ is a function which maps the exterior of $C$ onto the exterior of the unit circle $|w| = 1$ so that the points at infinity correspond to each other. In this case we have* (1.2) *valid not merely for $z$ in $B$ but also for $z$ interior to the largest curve $C_R$ which contains in its interior no singularity of $f_0(z)$—such a curve $C_R$ exists and is exterior to $C$—and the convergence in* (1.2) *is uniform on an arbitrary closed point set interior to this $C_R$.*

In the second theorem we shall be concerned with approximation on the boundary $C$ of a region $B$ by functions which satisfy certain auxiliary conditions at a finite number of preassigned points $P$ interior to $B$. These conditions are, at each point $P$, prescription of the value of the function together with the values of its first $k_P$ derivatives.

THEOREM B. *Let $B$ be an arbitrary simply connected region with the boundary $C$, and let the function $F(z)$ be analytic on $B+C$. Then there exists a function $F_0(z)$ analytic interior to $B$ satisfying arbitrary given auxiliary conditions, such that*

$$M = \overline{\text{bound}}\,|F(z) - F_0(z)|\,,\ z \text{ in } B,$$

*is less than*

$$\overline{\text{bound}}\,|F(z) - F_n(z)|\,,\ z \text{ in } B,$$

*for any other function $F_n(z)$ analytic interior to $B$ and satisfying the same auxiliary conditions. The expression $|F(z) - F_0(z)|$ approaches the same limit $M$ no matter how $z$ in $B$ approaches $C$. If the functions $F_n(z)$, $n = 1, 2, \cdots$, are analytic interior to $B$ and satisfy the given auxiliary conditions, then*

$$\lim_{n\to\infty} \left[\overline{\text{bound}}\,|F(z) - F_n(z)|\,,\ z \text{ in } B\right] = M$$

*implies*

$$\lim_{n\to\infty} F_n(z) = F_0(z)$$

*for $z$ in $B$, uniformly for $z$ on an arbitrary closed point set interior to $B$.*

*If the region $B$ is limited and if its boundary $C$ is also the boundary of an infinite region, the Tchebycheff polynomial $\pi_n(z)$ of degree $n$ for approximation to $F(z)$ on $C$ with the given auxiliary conditions is the polynomial of degree $n$ which satisfies the auxiliary conditions such that*

$$\max|F(z) - \pi_n(z)|\,,\ z \text{ in } B,$$

*is less than the corresponding expression formed for any other polynomial of degree $n$ and satisfying the auxiliary conditions; this Tchebycheff polynomial*

*$\pi_n(z)$ exists for $n$ sufficiently large and is unique. It is true if $F(z)$ has no singularities other than in the infinite region bounded by $C$, that*

$$\lim_{n\to\infty} [\max |F(z) - \pi_n(z)|, \ z \text{ in } B] = M,$$

*from which it follows that*

$$\lim_{n\to\infty} \pi_n(z) = F_0(z) \tag{1.3}$$

*for $z$ in $B$, uniformly on any closed point set interior to $B$.*

*If $B$ is the interior of a limited analytic Jordan curve $C$ and if $C_R$ has the same generic meaning as before, we have* (1.3) *valid not merely for $z$ in $B$ but also for $z$ interior to the largest curve $C_R$ which contains in its interior no singularity of $F_0(z)$—such a curve exists and is exterior to $C$—and the convergence in* (1.3) *is uniform on an arbitrary closed point set interior to this $C_R$.*

These two theorems and generalizations are to be proved by methods and results due to Carathéodory and Fejér, Gronwall, Schur, R. Nevanlinna, Carathéodory, and Tonelli. Some of the results of Theorem B have recently been obtained by Julia, for special auxiliary conditions of the form $F_n(0)=0$, $F_n'(0)=1$. This case is connected with the conformal mapping of the region $B$ onto a circle, as is the case of Theorem A for which $f(z)$ is of the form $1/z$.

F. Riesz has recently considered* an *integral* measure of approximation and states without giving any details that his methods can be used in studying approximation to rational functions by analytic functions. Those methods are not closely related to the ones that we use here.

The remainder of the present paper is devoted largely to proofs of the various parts of Theorems A and B, although some of our results are more detailed than is suggested by the statements just given, and others, particularly in §9, are considerably more general than Theorems A and B. In fact Theorems 14 and 15 together make a unit which includes both Theorems A and B as special cases.

**2. Existence and uniqueness of minimizing function for a circle.** The first part of Theorem B is now to be proved for the special case of a circle:

THEOREM 1. *Of the class of all functions, each of which is analytic for $|z|<1$ and each of which at $k$ given points $P$ interior to $C$: $|z|=1$ takes on preassigned values (the same for all functions of the class) for the function and its first $k_P$ derivatives, a unique function $F(z)$ has the least upper bound to its absolute value for $|z|<1$. This function $F(z)$ is a rational function of $z$ and of constant modulus $M$ on $C$.*

* Acta Mathematica, vol. 42 (1920), pp. 145–171.

This theorem was established by Carthéodory and Fejér* for the case of a single point $P$, new methods were introduced by Gronwall† and Schur,‡ the theorem was proved by Pick§ for arbitrary $k$ but $k_P=0$, and the general theorem is to be established by methods and results due to R. Nevanlinna.‖ Nevanlinna has derived necessary and sufficient conditions that a function satisfying auxiliary conditions of the kind we are considering be of modulus less than unity interior to $C$.

We indicate the proof¶ of Theorem 1. There exists at least one function analytic for $|z|\leqq 1$ which satisfies the auxiliary conditions at the points $P$ and which has a limited modulus for $|z|\leqq 1$; indeed, a polynomial can easily be written down which has this property. Hence there is a finite lower limit $M$ to the upper bounds of the moduli for $|z|<1$ of all functions $F(z)$ analytic for $|z|<1$ which satisfy the auxiliary conditions. *There exists a function $F(z)$ analytic for $|z|<1$ and satisfying the auxiliary conditions, whose modulus for $|z|<1$ has the upper bound $M$.* For there exist functions $F_n(z)$ analytic for $|z|<1$ and satisfying the auxiliary conditions, such that

$$\overline{\text{bound}}\,|F_n(z)| \leqq M+\frac{1}{n}, \text{ for } |z|<1,\ n=1,2,\cdots.$$

These functions $F_n(z)$ form a normal family in $|z|<1$, and from them can be extracted a subsequence which converges for $|z|<1$, uniformly for $|z|\leqq\rho<1$. Let the limit of this subsequence be denoted by $F(z)$. Then the function $F(z)$ also satisfies the auxiliary conditions, by the uniformity of the convergence, and we have in each point of $|z|<1$

$$|F(z)| \leqq M,$$

by the conditions on the $F_n(z)$.

The uniqueness of this function $F(z)$ analytic and of modulus not greater than $M$ for $|z|<1$ and satisfying the auxiliary conditions, follows readily from the results of Nevanlinna. We consider the new function** $F(z)/M$, with the obvious corresponding modifications to be made in the auxiliary conditions in the given points. If the function $F(z)$ is not unique, we have, in the

* Palermo Rendiconti, vol. 32 (1911), pp. 218–239.

† Annals of Mathematics, (2), vol. 14 (1912–13), pp. 72–80; vol. 16 (1914–15), pp. 77–81.

‡ Crelle's Journal, vol. 147 (1916), pp. 205–232; vol. 148 (1917), pp. 122–145.

§ Mathematische Annalen, vol. 77 (1915), pp. 7–23; Annales Academiae Scientiarum Fennicae, (A), vol. 15 (1921).

‖ Annales Academiae Scientiarum Fennicae, (A), vol. 13 (1920), pp. 1–71.

¶ Compare Gronwall, loc. cit.

** The case that the prescribed values of function and derivatives in the auxiliary conditions are all zero is trivial and henceforth excluded. In any other case we have $M\neq 0$.

notation of Nevanlinna (loc. cit., pp. 15, 36),

$$|\phi_n| < 1, \tag{2.1}$$

where our function $F(z)/M$ plays the rôle of Nevanlinna's function $z(x)$, and where $n+1$ is the sum of the $k_P$. We can then choose a particular $z^{(n)}(x)$ identically equal to $\phi_n$, a constant of absolute value less than unity. Then there exists a function $z(x)$ defined by Nevanlinna's recurrent equation (22) [if $k_P=0$ and by the corresponding equation if $k_P>0$], analytic for $|z|\leqq 1$ and satisfying the new auxiliary conditions. For this function it is true and follows directly from Nevanlinna's equation (22) and from our (2.1) that for $|z|\leqq 1$ we have

$$|z^{(n-1)}(x)| < \rho_{n-1} < 1, \qquad |z^{(n-2)}(x)| < \rho_{n-2} < 1, \cdots, |z(x)| < \rho < 1.$$

A special case of these inequalities, namely, $|\phi_i|<1$, $i=0, 1, \cdots, n-1$, is indeed necessary for the existence of the function $z(x)$. Thus we have, for some function $F(z)$ satisfying our original auxiliary conditions,

$$\left|\frac{F(z)}{M}\right| < \rho < 1, \text{ for } |z| \leqq 1,$$

which contradicts our definition of $M$.

We remark, and this follows from Nevanlinna's results, that our function $F(z)$ is of degree less than $n+1$, the sum of the $k_P$.

The effective determination of the number $M$ of Theorem 1 in any given case does not seem to be a simple matter. It will be noticed, however, that Nevanlinna's equation (pp. 15, 36)

$$|\phi_n(x_0, z_0; \cdots; x_n, z_n)| = 1, \tag{2.2}$$

where $n$ has the same significance as before, is an algebraic* equation which has $M$ for a root. Here the quantities $z_i$, etc., are not precisely the values which occur in the auxiliary conditions of Theorem 1, but are the quotients of those values by $M$. The condition on $M$ represented by (2.2) is necessary (see also below) that $M$ be the number of Theorem 1, because necessary for the uniqueness of Nevanlinna's function $z(x)$, but is not sufficient. The sufficiency arises only if Nevanlinna's inequalities

$$|\phi_0| < 1, \cdots, |\phi_{n-1}(x_0, z_0; \cdots; x_{n-1}, z_{n-1})| < 1, \tag{2.3}$$

where the $z_i$ are again the quantities of Theorem 1 divided by $M$, are also

* Equation (2.2) is not in its present form an algebraic equation, but becomes such if the various absolute values are suitably written and the corresponding equation simplified. There may be ambiguity, however, for the sign of a term expressed as an absolute value (such as $|M^2-\alpha|$) may depend on the value of $M$, which is itself determined by the equation.

satisfied. These inequalities are not implied by (2.2), as simple examples show.* But equation (2.2) has a unique positive root for which inequalities (2.3) are satisfied. For we shall reach a contradiction by assuming that there exist two such roots $M_1$ and $M_2$, with $M_1<M_2$. There exists (Nevanlinna, pp. 15, 36) a unique function $F_1(z)$ which takes on the prescribed auxiliary values and is of modulus not greater than $M_1$ in $C$ and also a unique function $F_2(z)$ which takes on the prescribed auxiliary values and is of modulus not greater than $M_2$ in $C$. This contradicts the obvious fact that the polynomial

$$F_1(z) + \epsilon \prod (z - \alpha_i)$$

where the $\alpha_i$ are the points $P$ and appear respectively $k_P$ times in the product, also satisfies the auxiliary conditions and for suitable sufficiently small $\epsilon$ is of modulus less than $M_2$ and yet not equal to $F_2(z)$.

Nevanlinna's condition (2.2), (2.3) is, as a matter of fact, not necessary and sufficient for the uniqueness of $z(x)$, and hence (2.2) is not in its present form a necessary condition on $M$. For it may occur that we have

$$(2.4) \qquad |\phi_0| < 1, \ |\phi_1| < 1, \cdots, \ |\phi_{i-1}| < 1, \ |\phi_i| = 1, \ i < n,$$
$$z_\nu^{(i)}(x_0, z_0; \cdots; x_{i-1}, z_{i-1}; x_\nu, z_\nu) = \phi_i, \ \nu = i+1, \ i+2, \cdots, n,$$

in which case also $z(x)$ exists and is unique. In this case the functions $z^{(i+1)}(x), \cdots, z^{(n)}(x)$ and quantities $|\phi_{i+1}|, \cdots, |\phi_n|$ take the form $0/0$ and are, properly speaking, not defined. But equation (2.2) can be written and transformed so that it is automatically satisfied also if (2.4) holds, and in this sense our previous statement is true, that (2.2) is an algebraic equation of which $M$ is a root.

**3. Continuity of the function $F(z)$.** We shall prove now, for later application, that *the function $F(z)$* (of Theorem 1) *varies continuously with the auxiliary conditions.* That is to say, if $\epsilon' > 0$ be given, there exists $\delta$ such that the inequalities

$$(3.1) \qquad |\alpha_i - \alpha_i'| < \delta, \ |\gamma_i - \gamma_i'| < \delta \qquad (i = 1, 2, \cdots, k)$$

imply that

$$(3.2) \qquad |F(z) - F'(z)| < \epsilon' \text{ for } |z| \leqq 1,$$

where $F(z)$ is the function of Theorem 1 for the auxiliary conditions

$$(3.3) \qquad F(\alpha_i) = \gamma_i \qquad (i = 1, 2, \cdots, k),$$

and $F'(z)$ is the function of Theorem 1 for the auxiliary conditions

$$(3.4) \qquad F'(\alpha_i') = \gamma_i' \qquad (i = 1, 2, \cdots, k).$$

* For instance, $x_0=0, x_1=1/2, z_0=2, z_1=5/4$.

Here the values $\alpha_i$ and $\gamma_i$ are considered fixed, while $\alpha_i'$ and $\gamma_i'$ are arbitrary, subject merely to the restrictions (3.1). The number $\delta$ will then depend on $\epsilon'$ and on the $\alpha_i$ and $\gamma_i$, but not on the $\alpha_i'$ and $\gamma_i'$. It is reasonable to suppose that $F(z)$ varies continuously with the auxiliary conditions, for $F(z)$ is determined from those auxiliary conditions entirely by algebraic processes. We give, however, a formal proof.

As a preliminary proposition we shall prove that *the absolute value $M$ of $F(z)$ for $|z|=1$ varies continuously with the auxiliary conditions.* That is, if $\epsilon>0$ be given, there exists $\delta$ such that inequalities (3.1) imply

$$M-\epsilon<M'<M+\epsilon, \tag{3.5}$$

where $M'$ is the absolute value of $F'(z)$ for $|z|=1$. It seems reasonable too that $M$ should be a continuous function of the auxiliary conditions, for $M$ is a root of an algebraic equation whose coefficients depend rationally on those conditions. But this algebraic equation may depend, as we have already suggested, to some extent on the value of $M$ itself, so we prefer to give a detailed proof based on other considerations.

It is possible so to determine $\delta$ that whenever we have

$$|\alpha_i-a_i|<\delta_1,\quad |\gamma_i-c_i|<2\delta, \tag{3.6}$$

where $2\delta_1$ is less than every $|\alpha_i-\alpha_j|$, $i\neq j$, then there exists* a polynomial $p(z)$ which takes on the values $c_i-\gamma_i$ at the points $a_i$ and is such that $|p(z)|\leqq\epsilon$ for $|z|\leqq 1$. The numbers $\alpha_i$ and $\gamma_i$ are as before considered fixed, determined by the auxiliary conditions (3.3), while the $a_i$ and $c_i$ are subject only to the restriction of inequalities (3.6). All functions analytic and of absolute value not greater than $M+\epsilon$ for $|z|\leqq 1$ are equicontinuous for $|z|\leqq\rho<1$. We choose $\rho$ greater than every $|\alpha_i|$, and then still further restrict $\delta_1$ in (3.6) if necessary so that $\delta_1+|\alpha_i|<\rho$ and so that functional values at two arbitrary points in $|z|\leqq\rho$ whose distance apart is less than $\delta_1$, of an arbitrary function analytic and of modulus not greater than $M+\epsilon$ in $|z|\leqq 1$, differ by less than $\delta$.

We shall now prove (3.5) under the assumption that the $\delta$ of (3.1) is the smaller of the two quantities $\delta_1$ and $\delta$ of (3.6). Let us consider the polynomial $p(z)$ such that

$$p(\alpha_i')=F'(\alpha_i')-F(\alpha_i)+[F(\alpha_i)-F(\alpha_i')]=\gamma_i'-\gamma_i+[F(\alpha_i)-F(\alpha_i')],$$

which is in absolute value less than $2\delta$. We have $|p(z)|\leqq\epsilon$, for $|z|\leqq 1$, and

* Walsh, these Transactions, vol. 30 (1928), pp. 307–332. See the proof of the Lemma, p. 319. We are using in the present paper essentially a generalization of that Lemma, where the functional values $c_i-\gamma_i$ of the polynomial concerned are restricted only by the second of inequalities (3.6), and the points $a_i$ at which those values are taken on are restricted only by the first of those inequalities.

hence $|F(z)+p(z)| \leqq M+\epsilon$, for $|z| \leqq 1$, where $F(\alpha_i')+p(\alpha_i')=F'(\alpha_i')=\gamma_i'$. That is, we have exhibited a function $F_1(z)=F(z)+p(z)$ analytic for $|z| \leqq 1$ whose functional values are $\gamma_i'$ for $z=\alpha_i'$, and whose modulus for $|z| \leqq 1$ is not greater than $M+\epsilon$. Thus we have proved $M' \leqq M+\epsilon$.

We determine next a new polynomial $p'(z)$ such that

$$\begin{aligned} p'(\alpha_i) &= F(\alpha_i) - F'(\alpha_i') + [F'(\alpha_i') - F'(\alpha_i)] \\ &= \gamma_i - \gamma_i' + [F'(\alpha_i') - F'(\alpha_i)], \end{aligned}$$

which is in absolute value less than $2\delta$; we have already shown $|F'(z)| \leqq M+\epsilon$ for $|z| \leqq 1$, so the square bracket is in absolute value less than $\delta$. It follows that $|p'(z)| \leqq \epsilon$ for $|z| \leqq 1$, and hence that $|F'(z)+p'(z)| \leqq M'+\epsilon$ for $|z| \leqq 1$, where $F'(\alpha_i)+p'(\alpha_i)=F(\alpha_i)=\gamma_i$. That is, we have exhibited a function $F_2(z)=F'(z)+p'(z)$ analytic for $|z| \leqq 1$ whose functional values are $\gamma_i$ for $z=\alpha_i$, and whose absolute value for $|z| \leqq 1$ is not greater than $M'+\epsilon$. Thus we have proved $M \leqq M'+\epsilon$, and the preliminary proposition (3.5) is established.

To be sure, we have given the proof only for the case $k_P=0$, but this is simply a matter of convenience in exposition. The modifications necessary for arbitrary $k_P$ involve inequalities of form (3.1) also for the derivatives of $F(z)$ at the points $\alpha_i$, but are fairly obvious and hence are left to the reader. This remark applies also below.

We can now easily prove the desired continuity of the function $F(z)$. For the formulas given by Nevanlinna (i.e. his (22)), suitably modified for functions of modulus not greater than $M$ instead of not greater than unity, give us $F(z)$ explicitly in terms of $M$ and of the quantities involved in the auxiliary conditions. First we determine $\phi_n$ in terms of these quantities and then express $z(x)$ (modified) in terms of $\phi_n$ and these quantities. These formulas for $F(z)$ in terms of all these quantities are indeed valid even if we have Nevanlinna's equation (30) for some $i<n$. In the modified (22) for instance, we have

$$\frac{z^{(i)}(x)}{M} = \frac{\xi_i \dfrac{z^{(i+1)}(x)}{M} + \dfrac{\phi_i}{M}}{1 + \dfrac{\bar{\phi}_i \xi_i z^{(i+1)}(x)}{M^2}};$$

if $z^{(i+1)}(x)$ approaches the limit $\phi_i$, which must occur if we are to have (30) and (30′) fulfilled, then we have $|\phi_i|=M$, and hence we find $\lim z^{(i)}(x)=\phi_i$, uniformly throughout the unit circle. The functions $F(z)$ involved are in every case seen by inspection to be continuous functions of $M$ and of the auxiliary conditions, for $z$ on and within $C$, and our theorem is established.

It may be remarked that a simpler proof of this theorem may be given without recourse to Nevanlinna's formulas if we are content to establish the continuous variation of $F(z)$ merely for $|z| \leqq \rho < 1$. This new proof depends on the continuity of $M$ considered as a function of the initial conditions and on the uniform continuity for $|z| \leqq \rho < 1$ of functions analytic and of modulus not greater than $M+\epsilon$ for $|z| \leqq 1$. The proof is in spirit closely related to the proof of Theorem 2 below and can be easily supplied by the reader.

4. **Approximation on a circle.** In the notation of Theorem 1, we can prove at once another part of Theorem B:

THEOREM 2. *Let the functions $F_n(z)$ analytic for $|z| < 1$ take on, at the $k$ given points $P$ interior to $C: |z| = 1$, the prescribed values for the functions and for their first $k_P$ derivatives. Then*

$$\lim_{n \to \infty} [\overline{\text{bound}}\, F_n(z),\ |z| < 1] = M \tag{4.1}$$

*implies*

$$\lim_{n \to \infty} F_n(z) = F(z) \tag{4.2}$$

*for $|z| < 1$, uniformly for $|z| \leqq \rho$, where $\rho$ is an arbitrary number less than unity. Here the function $F(z)$ is the function of Theorem 1 corresponding to the prescribed auxiliary conditions.*

Interior to $C$ the functions $F_n(z)$ are uniformly bounded and hence form a normal family there. Then from any infinite set of the $F_n(z)$ can be extracted a subsequence which converges for $|z| < 1$, uniformly for $|z| \leqq \rho < 1$. The limit of such a subsequence is, by (4.1), a function whose modulus in $C$ is not greater than $M$. This limit function is analytic for $|z| < 1$ and by the uniformity of the convergence takes on at the $k$ given points $P$ the prescribed values for itself and for its first $k_P$ derivatives, and hence by Theorem 1 coincides with $F(z)$. That is to say, from every infinite set of the $F_n(z)$ can be extracted a subsequence whose limit is $F(z)$, so the original sequence $F_n(z)$ has the limit $F(z)$ for $|z| < 1$, and being uniformly bounded converges to the limit uniformly for $|z| \leqq \rho < 1$.

The simplest cases of Theorem 2 are the cases where the prescribed auxiliary conditions are of the form

$$F_n(0) = 0,\quad F_n'(0) = 0, \cdots, F_n^{(k)}(0) = 1;$$

here Theorems 1 and 2 can be readily proved by elementary means, by studying $k!F_n(z)/z^k$, and of course it turns out that the corresponding function $F(z)$ of Theorem 1 is $z^k/k!$.

A detailed study of the behavior of the functions $F_n(z)$ of Theorem 2 on the circumference $C$ is beyond the scope of the present paper,* but it is perhaps of interest to give an elementary example indicating that the convergence in (4.2) need not be uniform for $|z|<1$. Our example shows in fact that $\lim_{n\to\infty}F_n(z)$ need not be continuous for $|z|\leqq 1$ even when the $F_n(z)$ are analytic for $|z|\leqq 1$. We set

$$F_n(z) = e^{1/[n(z-1)-1]+1/(n+1)},$$

which is analytic for $|z|\leqq 1$ and has the value unity at the origin. This prescription $(F_n(0)=1)$ is taken as the sole auxiliary condition on the functions $F_n(z)$, and the corresponding value of $M$ is unity. It is, moreover, clear that we have

$$\lim_{n\to\infty} F_n(z) = \begin{cases} 1, & z \neq 1, \\ \dfrac{1}{e}, & z = 1, \end{cases}$$

so that the limit function is discontinuous† not merely for $|z|\leqq 1$ but also for $|z|=1$. The absolute value of $F_n(z)$ is the exponential function of the real part of

$$\frac{1}{n(z-1)-1}+\frac{1}{n+1},$$

that is, the exponential function of

$$\theta(x,y) = \frac{n(x-1)-1}{(nx-n-1)^2+n^2y^2}+\frac{1}{n+1}.$$

On the circle $|z|=1$ we have

$$\theta(x,y) = \frac{n(x-1)-1}{1-2(n^2+n)(x-1)}+\frac{1}{n+1},$$

whose maximum value occurs for $x=-1$:

$$\theta(-1,0) = \frac{1}{n+1}-\frac{1}{2n+1}.$$

This expression approaches zero with $1/n$, so (4.1) is indeed satisfied.

---

* See also §11.

† It is not difficult to show that in Theorem 2

$$\lim_{n\to\infty} |F_n(z)| < M$$

is impossible on a point set $C$ of Lebesgue measure greater than zero.

The particular sequence $F_n(z)$ may be easily replaced by a sequence of polynomials in $z$ which has the same convergence properties for $|z| \leqq 1$. In fact, we can prove that *if $\{F_n(z)\}$ is an arbitrary sequence of functions analytic interior to a Jordan curve $C$ and continuous in the corresponding closed region $\bar{C}$, then there exists a sequence $\{P_n(z)\}$ of polynomials in $z$ converging in $\bar{C}$ whenever the given sequence converges and to the same sum, diverging in $\bar{C}$ whenever the given sequence diverges, converging uniformly on any point sets belonging to $\bar{C}$ on which the given sequence converges uniformly, and such that we have $P_n(z) = F_n(z)$ in an arbitrary finite number of preassigned points $\alpha_i$ of $\bar{C}$.* If the latter points are interior to $C$, we may require equality also of an arbitrary finite number of derivatives

$$P_n^{(k)}(\alpha_i) = F_n^{(k)}(\alpha_i) \qquad (k = 0, 1, \cdots, k_i)$$

at the preassigned points of $\bar{C}$.

There exist polynomials $P_n(z)$ such that we have

$$| P_n(z) - F_n(z) | < \frac{1}{n},$$

for $z$ in $\bar{C}$, and where we may choose $P_n^{(k)}(\alpha_i) = F_n^{(k)}(\alpha_i)$ as described.* The sequence $\{P_n(z)\}$ has all the properties required.

A simple transformation of the results considered in Theorem 2 will yield new results that we desire, parts of Theorem A.

THEOREM 3. *Let $r(z)$ be a rational function of $z$ with all of its singularities interior to $C: |z| = 1$. Of all functions $f(z)$ analytic for $|z| < 1$, there exists a unique function such that*

$$\overline{\operatorname{bound}}_{\substack{|z|\to 1 \\ |z|<1}} | r(z) - f(z) | \tag{4.3}$$

*is least; this function $f(z)$ is rational, and $r(z) - f(z)$ is of constant modulus $M$ on $C$.*

*If the functions $f_n(z)$ are analytic for $|z| < 1$, then*

$$\lim_{n\to\infty} \left[ \overline{\operatorname{bound}}_{\substack{|z|\to 1 \\ |z|<1}} | r(z) - f_n(z) | \right] = M \tag{4.4}$$

*implies*

$$\lim_{n\to\infty} f_n(z) = f(z)$$

*for $|z| < 1$, uniformly for $|z| < \rho$, provided merely $\rho < 1$.*

---

* Compare the proof of Theorem 7 below.

Let the given function be

$$r(z) = \frac{a_0 z^\nu + a_1 z^{\nu-1} + \cdots + a_\nu}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_\mu)} = A_0 + \frac{A_1}{z-\alpha_1} + \frac{A_2}{z-\alpha_2} + \cdots + \frac{A_\mu}{z-\alpha_\mu}, \tag{4.5}$$

where all the $\alpha_i$ lie interior to the circle $C$. If the function $f(z)$ is analytic within $C$, continuous on and within $C$, then the function*

$$F(z) \equiv [r(z) - f(z)]\frac{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_\mu)}{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)\cdots(1-\bar{\alpha}_\mu z)} \tag{4.6}$$

is likewise analytic within $C$, continuous on and within $C$, and has on $C$ the same maximum absolute value† as $r(z)-f(z)$. In any case if $f(z)$ is analytic within $C$ (for the present $f(z)$ is subject to no other restriction), so also is $F(z)$, and it is true that

$$\overline{\operatorname{bound}_{|z|<1}} F(z) = \overline{\operatorname{bound}_{\substack{|z|\to 1\\ |z|<1}}} [r(z) - f(z)]. \tag{4.7}$$

In the point $\alpha_i$ the function $F(z)$ has the value

$$\frac{a_0\alpha_i^\nu + a_1\alpha_i^{\nu-1} + \cdots + a_\nu}{(1-\bar{\alpha}_1\alpha_i)(1-\bar{\alpha}_2\alpha_i)\cdots(1-\bar{\alpha}_\mu\alpha_i)} = A_i\frac{(\alpha_i-\alpha_1)(\alpha_i-\alpha_2)\cdots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\cdots(\alpha_i-\alpha_\mu)}{(1-\bar{\alpha}_1\alpha_i)(1-\bar{\alpha}_2\alpha_i)\cdots(1-\bar{\alpha}_\mu\alpha_i)}, \tag{4.8}$$

which depends only on $r(z)$ and not in any way on $f(z)$. If two or more of the $\alpha_i$ in $r(z)$ are equal, that is to say, if $r(z)$ has poles of order higher than the first, then formula (4.5) requires some revision, and contains denominators of degree higher than the first. In this case not merely is the value of $F(\alpha_i)$ determined by (4.6) and independent of $f(z)$, but also the value of the first and possibly higher derivatives of $F(z)$ at $z=\alpha_i$, depending on the number of the $\alpha_i$ which coalesce. The reader can easily make the necessary modifications in the formulas given, to conform to this new situation. In every case the auxiliary conditions (4.8) are of precisely the kind we have been studying in Theorems 1 and 2.

* The case that $r(z)$ is a constant is trivial and henceforth excluded.

† For $z$ on $C$ we have $z\bar{z}=1$, so the expression

$$\frac{z-\alpha_1}{1-\bar{\alpha}_1 z} = \frac{z\bar{z}-\alpha_1\bar{z}}{\bar{z}-\bar{\alpha}_1 z\bar{z}} = \frac{1-\alpha_1\bar{z}}{\bar{z}-\bar{\alpha}_1}$$

equals the reciprocal of its conjugate and hence has the absolute value unity.

Reciprocally, let the function $F(z)$ analytic interior to $C$ be given, which in the points $\alpha_i$ takes on the values (4.8). Let us write $F(z)$ in the form indicated by (4.6), where $r(z)$ is given by (4.5) as before. Then the function $f(z)$ defined by (4.6) is analytic interior to $C$ and is continuous on and within $C$ if $F(z)$ has that property; the maximum absolute value of $F(z)$ on $C$ is the same as that of $r(z)-f(z)$. In every case (4.7) is true. Here the formulas (4.5), (4.6), (4.8) are of course to be somewhat modified if $r(z)$ has multiple poles, but the modifications present no difficulty and are left to the reader.

The first part of Theorem 3 now follows immediately from Theorem 1, for we have transformed the situation of Theorem 3 into an equivalent situation in Theorem 1. The second part of Theorem 3 follows without difficulty from Theorem 2. In fact let us replace $f(z)$ in (4.6) by $f_n(z)$ and $F(z)$ by $F_n(z)$; this new equation

$$F_n(z) \equiv [r(z) - f_n(z)]\frac{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_\mu)}{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)\cdots(1-\bar{\alpha}_\mu z)}$$

is the definition of $F_n(z)$, whereas equation (4.6) is now considered to define the function $f(z)$ of Theorem 3 in terms of the function $F(z)$ of Theorem 1 for the auxiliary conditions (4.8). Then (4.4) implies

$$\lim_{n\to\infty} [\overline{\text{bound}}_{|z|<1} F_n(z)] = M,$$

and hence by Theorem 2 implies

$$\lim_{n\to\infty} F_n(z) = F(z)$$

for $|z|<1$, uniformly for $|z|\leqq\rho$, where $F(z)$ is the function of Theorem 1. This equation implies in turn

$$\lim_{n\to\infty} f_n(z) = f(z), \tag{4.9}$$

where $f(z)$ is defined by (4.6) in terms of the function $F(z)$ of Theorem 1, and (4.9) holds for $|z|<1$, uniformly for $|z|\leqq\rho<1$, except in the neighborhoods of the points $\alpha_i$. But the functions $f_n(z), f(z)$ are analytic in the neighborhoods of those points as well as at the points $\alpha_i$, and equation (4.9), holding uniformly *on* small circles interior to $C$ surrounding the respective points $\alpha_i$, implies that (4.9) holds also *interior* to those small circles, so Theorem 3 is completely established.

It is not hard to obtain some information about the degree of the function $f(z)$ of Theorem 3. We find from (4.6),

$$f(z) \equiv r(z) - \frac{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)\cdots(1-\bar{\alpha}_\mu z)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_\mu)}F(z).$$

The degree of $F(z)$ is less than $\mu$, and the degree of $r(z)$ is precisely $\mu$. When the two fractions in the right-hand member are reduced to a common denominator, the new fraction is of degree less than $2\mu$. But the function $f(z)$ has no singularities interior to $C$, so the factors $z-\alpha_i$ are common to numerator and denominator, and the actual degree of $f(z)$ is less than $\mu$.

The simplest case of Theorem 3 is for approximation to the function

$$r(z) = \frac{1}{z^k}, \quad k > 0.$$

Here the proof of the theorem can be given by elementary methods, and of course the function $f(z)$ is identically zero.

**5. Some inequalities.** It is reasonable to suppose that there exist inequalities, more explicit than Theorem 2, which give an upper limit for $|F_n(z)-F(z)|$ in terms of $\{[\overline{\text{bound}}\, F_n(z),\ |z|<1]-M\}$, indicating the fact that approach to zero of the latter quantity implies approach to zero of the former, if $z$ is properly restricted. Theorem 2 itself can be used to establish the existence of such inequalities, but we use the direct method based on the formulas of Nevanlinna. We shall treat in full detail only a special case, but shall indicate later the modifications necessary to treat the general case.

We state for reference what is essentially a particular case of the inequalities we propose to establish:

CARATHÉODORY'S LEMMA.* *If the function $F(z)$ is analytic and in absolute value not greater than $N$ for $|z|<1$, then we have*

$$|F(z) - F(0)| \leqq |z| \frac{N^2 - |F(0)|^2}{N - |F(0)| \cdot |z|}. \tag{5.1}$$

If we set $N = |F(0)| + \nu$, this inequality can be written

$$|F(z) - F(0)| \leqq |z| \frac{\nu[\nu + 2|F(0)|]}{\nu + |F(0)|(1 - |z|)}. \tag{5.2}$$

It follows from (5.2) that if we have a sequence of functions $F_n(z)$ having the common value $F(0)$ for $z=0$, and if we have for $|z|<1$

$$|F_n(z)| \leqq |F(0)| + \nu_n, \qquad \lim_{n\to\infty} \nu_n = 0,$$

then we have $\lim_{n\to\infty} F_n(z) = F(0)$ uniformly for $|z| \leqq \rho < 1$. This is a special case of Theorem 2, which together with the corresponding special case of

* Mathematische Annalen, vol. 72 (1912), p. 107. Pólya und Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. I, pp. 139, 323. The Lemma there has the hypothesis $|F(z)|<N$, instead of $|F(z)| \leqq N$, but since it holds in the one case it holds also in the other.

Theorem 1 can easily be proved without the aid of Carathéodory's Lemma. Nevertheless (5.2) is more explicit and for some purposes more useful, as we shall see, than Theorem 2.

We treat now the special situation of Theorem 2 where $k=1$, and $P$ is the origin. The function $F(z)$ of Theorem 2 may be defined by the recurrent process of setting (Nevanlinna, loc. cit. p. 37; the formulas for this special case are those of Gronwall and Schur)

$$(5.3)\quad \phi_l(z) = \frac{M^2}{z}\,\frac{\phi_{l-1}(z) - \phi_{l-1}(0)}{M^2 - \bar{\phi}_{l-1}(0)\phi_{l-1}(z)}, \quad l = 1,2,\cdots,p, \quad \phi_0(z) \equiv F(z),$$

where $\phi_p(z)$ is a constant of modulus $M$ and where $p$ is the $k_0$ of Theorem 1.* In exceptional cases it may occur that $\phi_q(z)$ is a constant (necessarily of modulus $M$) for some $q<p$. This involves a change in the notation but no essential change in the reasoning; the corresponding modifications here and below are left to the reader.

If we have now the functions $F_n(z)$ of Theorem 2 with the present auxiliary conditions, we set likewise

$$(5.4)\qquad \phi_{nl}(z) = \frac{M^2}{z}\,\frac{\phi_{n,l-1}(z) - \phi_{n,l-1}(0)}{M^2 - \bar{\phi}_{n,l-1}(0)\phi_{n,l-1}(z)}, \quad \phi_{n0}(z) = F_n(z).$$

It follows that we have

$$\phi_l(0) = \phi_{nl}(0), \quad l = 1,2,\cdots,p;\ n = 1,2,\cdots;$$
$$|\phi_l(z)| \leqq M \text{ for } |z|<1, \quad l = 0,1,\cdots,p.$$

Direct computation yields

$$(5.5)\quad \phi_{nl}(z) - \phi_l(z) = \frac{M^2}{z}\,\frac{[\phi_{l,n-1}(z) - \phi_{l-1}(z)][M^2 - \phi_{l-1}(0)\bar{\phi}_{l-1}(0)]}{[M^2 - \bar{\phi}_{l-1}(0)\phi_{n,l-1}(z)][M^2 - \bar{\phi}_{l-1}(0)\phi_{l-1}(z)]}.$$

Under our present hypothesis we have $|\phi_l(0)|<M$, for $l=0,1,\cdots,p-1$, so all of the factors $M^2-\phi_{l-1}(0)\,\bar{\phi}_{l-1}(0)$, $M^2-\bar{\phi}_{l-1}(0)\,\phi_{l-1}(z)$, $M^2-\bar{\phi}_{l-1}(0)\cdot\phi_{n,l-1}(z)$ are bounded from zero, except possibly the last, and this is bounded from zero if $n$ is sufficiently great. We remark too that by virtue of the equation $\phi_{n,l-1}(0)=\phi_{l-1}(0)$, the function

$$\frac{\phi_{n,l-1}(z) - \phi_{l-1}(z)}{z}$$

is analytic for $|z|<1$, and hence has the upper limit of its modulus in $|z|<1$ not for $z=0$ but for $|z|\to 1$.

* We write, here and below, $\bar{\phi}_k(z)$ to denote the conjugate of $\phi_k(z)$.

Equations (5.5) for $l=1, 2, \cdots, p$ are to be used twice in deriving the inequalities we desire. Moreover it is the nature of the inequalities rather than their precise form that interests us, although the precise form can be readily obtained by the reader. We introduce the notation, under the hypothesis of Theorem 2,

$$M + \epsilon_n = \overline{\text{bound}}\,|F_n(z)| \quad \text{for} \quad |z| < 1, \tag{5.6}$$

from which it follows that $\lim_{n\to\infty}\epsilon_n = 0$.

By studying the transformation involved in (5.4), it is easily shown that the inequality $|\phi_{n,l-1}(z)| \leqq m_{n,l-1}$ implies

$$|\phi_{nl}(z)| \leqq M^2 \frac{m_{n,l-1} - |\phi_{l-1}(0)|}{M^2 - m_{n,l-1}|\phi_{l-1}(0)|}, \quad \text{for } |z| < 1, \tag{5.7}$$

if $m_{n,l-1}$ is sufficiently near to $M$, and we apply this formula to each of the functions $\phi_{n1}(z), \phi_{n2}(z), \cdots, \phi_{np}(z)$ in turn. It is obvious that the right-hand member of (5.7) approaches $M$, as $m_{n,l-1}$ approaches $M$, and also that the derivative of the right-hand member with respect to $m_{n,l-1}$ for the value $m_{n,l-1}=M$ is finite, from which it follows that we have

$$|\phi_{nl}(z)| \leqq M + M_l(m_{n,l-1} - M), \text{ for } |z| < 1,$$

where $M_l$ is a suitably chosen number independent of $n$.

We use this inequality for $l=1, 2, \cdots, p$ in succession, setting $m_{nl} = M + M_l(m_{n,l-1} - M)$, $m_{n0} = M + \epsilon_n$, and derive

$$|\phi_{np}(z)| \leqq M + M'\epsilon_n$$

where $M'$ is a suitably chosen number independent of $n$. Finally Carathéodory's Lemma informs us that we have

$$|\phi_{np}(z) - \phi_p(z)| \leqq M''\epsilon_n \text{ for } |z| \leqq \rho < 1, \tag{5.8}$$

where $M''$ depends on $\rho$ but not on $n$; for it is true that $\phi_p(z) \equiv \phi_p(0)$. Immediate use of equation (5.5) for $l=p, p-1, \cdots, 1$ then gives inequalities similar to (5.8) where $p$ is replaced successively by $p-1, p-2, \cdots, 0$, and finally we have

$$|F_n(z) - F(z)| \leqq M_1'\epsilon_n, \text{ for } |z| \leqq \rho < 1, \tag{5.9}$$

where $M_1'$ depends on $\rho$ but not on $\epsilon_n$. All of these inequalities have been established only if $\epsilon_n$ is sufficiently small, but do hold under such restrictions, provided merely that $\epsilon_n$ is defined by (5.6).

We have given the derivation of (5.9) not for the most general situation of Theorem 2, but under the assumption that only the values of $F_n(z)$ and

of its first $p$ derivatives are assigned at the origin. In the general case, the formulas given are to be used, and followed by others of the form

$$\psi_1(z) = M^2 \frac{\psi_0(z) - \psi_0(z_0)}{M^2 - \bar{\psi}_0(z_0)\psi_0(z)} : \frac{z - z_0}{1 - \bar{z}_0 z}, \tag{5.10}$$

$$\psi_{n1}(z) = M^2 \frac{\psi_{n0}(z) - \psi_{n0}(z_0)}{M^2 - \bar{\psi}_{n0}(z_0)\psi_{n0}(z)} : \frac{z - z_0}{1 - \bar{z}_0 z}, \tag{5.11}$$

which correspond to (5.3) and (5.4) respectively. We have

$$\psi_0(z) = F(z) \text{ or } \phi_p(0), \quad \psi_{n0}(z) = F_n(z) \text{ or } \phi_{np}(z),$$

and of course we have $\psi_0(z_0) = \psi_{n0}(z_0)$. We compute now the analogue of (5.5):

$$\psi_{n1}(z) - \psi_1(z) = \frac{M^2}{\dfrac{z - z_0}{1 - \bar{z}_0 z}} \frac{[\psi_{n0}(z) - \psi_0(z)][M^2 - \psi_0(z_0)\bar{\psi}_0(z_0)]}{[M^2 - \bar{\psi}_0(z_0)\psi_0(z)][M^2 - \bar{\psi}_{n0}(z_0)\psi_{n0}(z)]}.$$

By reasoning similar to that already given we now obtain (5.9) as before, provided that $\epsilon_n$ is sufficiently small. This holds whether we have a single application of formulas (5.10) and (5.11), or repeated application, and also with or without further repeated application of formulas of the type (5.3) and (5.4).

This gives us indeed the most general situation under Theorem 2, for we can first transform an arbitrary point $P$ to the origin, and use (5.3) and (5.4) to account for prescribed values of function and derivatives at $P$; then we can transform a second point $P$ to the origin and account for the prescribed values at this new point by means of (5.10) and (5.11) followed if necessary by (5.3) and (5.4), and so on. The inequalities we are using and deriving are in character unaltered by transformations of the types used.

Inequality (5.9) contains the essence of Theorem 2.

Julia has recently used* a special case of (5.9) in a study to which more detailed reference will be made later.

The discussion just given can be applied with ease by the methods of §4 to the situation of Theorem 3. We introduce the notation

$$M + \epsilon_n = \overline{\text{bound}} \mid r(z) - f_n(z) \mid \text{ for } \mid z \mid \to 1, \mid z \mid < 1.$$

The functions $F_n(z)$ are defined as in (4.6):

$$F_n(z) \equiv [r(z) - f_n(z)] \frac{(z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_\mu)}{(1 - \bar{\alpha}_1 z)(1 - \bar{\alpha}_2 z) \cdots (1 - \bar{\alpha}_\mu z)}; \tag{5.12}$$

* Annales de l'École Normale Supérieure, (3), vol. 44 (1927), pp. 289–316.

thus (5.6) may equally well be used as a definition of $\epsilon_n$. For $\epsilon_n$ sufficiently small, we have shown (5.9) to be valid. Direct computation by means of (4.6) and (5.12) then implies

$$|f_n(z) - f(z)| \leqq M_2'\epsilon_n \text{ for } |z| \leqq \rho < 1. \tag{5.13}$$

Inequality (5.13) follows from (4.6) and (5.12) first for $|z| = \rho$, where $|\alpha_i| < \rho$, but holding on the circle $|z| = \rho$ holds also within this circle as well.

**6. Tchebycheff polynomials.** A special sequence of approximating functions $\{f_n(z)\}$ or $\{F_n(z)\}$ which has interesting properties will now be studied.

If $\phi(z)$ is a continuous function defined on the closed limited point set $C$ of the $z$-plane, then there exists* a unique polynomial $\pi_n(z)$ of degree† $n$ such that

$$\max |\phi(z) - \pi_n(z)|, \quad z \text{ on } C, \tag{6.1}$$

is less than the corresponding expression

$$\max |\phi(z) - p_n(z)|, \quad z \text{ on } C, \tag{6.2}$$

for any other polynomial $p_n(z)$ of degree $n$. The polynomial $\pi_n(z)$ is called the *Tchebycheff polynomial* of degree $n$ for approximation to $\phi(z)$ on $C$, and has many important and well known properties. The following theorem, whose various parts are due to various writers,‡ will yield us the results on Tchebycheff polynomials contained in Theorems A and B.

THEOREM 4. *Let $C$ be a closed limited point set (not a single point) of the $z$-plane whose complementary set with respect to the entire plane is simply connected. Denote by $C_R$ the image in the $z$-plane of the circle $|w| = R > 1$ when the set complementary to $C$ is mapped onto the exterior of the unit circle in the $w$-plane so that the points at infinity correspond to each other. Then if $f(z)$ is analytic in the closed interior of $C_R$, there exist polynomials $p_n(z)$ of respective degrees $n = 0, 1, 2, \cdots$, such that we have*

$$|f(z) - p_n(z)| \leqq \frac{K}{R^n}, \quad z \text{ on } C, \quad n = 0, 1, 2, \cdots, \tag{6.3}$$

*where $K$ is independent of $n$ and $z$.*

---

* See §10 below. The theorem (due to Tonelli) assumes that $C$ contains at least $n+1$ distinct points.

† That is, $\pi_n(z)$ can be written in the form $a_0z^n + a_1z^{n-1} + \cdots + a_n$.

‡ References are given by Walsh, Münchner Berichte, 1926, pp. 223–229. In this theorem and below we tacitly assume that if the function $f(z)$ is not defined on the entire point set considered, then the function is to be defined on the new points by analytic extension, or (what amounts to the same thing) by the convergent sequence of polynomials.

*If the polynomials $p_n(z)$ are given so that* (6.3) *is valid for $z$ on $C$, where $f(z)$ is defined merely on $C$, then the sequence $\{p_n(z)\}$ converges interior to $C_R$, $R>1$, uniformly on any closed point set interior to $C_R$, and $f(z)$ is analytic interior to $C_R$. Moreover, an inequality of the form*

$$| f(z) - p_n(z) | \leqq K_1\left(\frac{R_1}{R}\right)^n \qquad (n = 0,1,2, \cdots)$$

*is valid for $z$ on and within an arbitrary $C_{R_1}$, $R_1<R$.*

In particular, if $f(z)$ is an entire function, the polynomials $p_n(z)$ exist so that for an arbitrary $R>1$ and for suitable $K$ (depending on $R$), the inequality (6.3) holds, and the sequence $\{p_n(z)\}$ converges over the entire plane, uniformly on any closed point set.

If in Theorem 4 the polynomials $p_n(z)$ are given so that (6.3) holds merely *for $n$ sufficiently large*, say $n \geqq N$, then arbitrary polynomials can be chosen as the $p_n(z)$, $n<N$, if those polynomials are not already defined, and for a suitably modified $K$ the inequality (6.3) will hold for all values of $n$. That is, it is still true that the sequence $\{p_n(z)\}$ converges interior to $C_R$, uniformly on an arbitrary point set interior to $C_R$.

With the same notation as in Theorem 4, we can now prove* easily

THEOREM 5. *If $f(z)$ is analytic interior to $C_R$ and if $\pi_n(z)$ is the Tchebycheff polynomial for approximation to $f(z)$ on $C$, then we have*

$$| f(z) - \pi_n(z) | \leqq \frac{K_1}{R_1^n}, \quad z \text{ on } C, \quad R_1 < R, \tag{6.4}$$

*and the sequence $\pi_n(z)$ converges throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

Theorem 5 is a direct consequence of Theorem 4. For if an arbitrary $R_1<R$ is given, there exist polynomials $p_n(z)$ of respective degrees $n=0, 1, 2, \cdots$, such that we have

$$| f(z) - p_n(z) | \leqq \frac{K_1}{R_1^n}, \quad z \text{ on } C.$$

Since this inequality is valid for the polynomials $p_n(z)$ it is likewise valid for the Tchebycheff polynomials $\pi_n(z)$, so that (6.4) holds. By virtue of Theorem 4, Theorem 5 now follows as stated. We note that if $f(z)$ has a singularity

* Theorem 5 was proved by Faber in the special case that $C$ is a region bounded by an analytic Jordan curve, by a method quite different from the present one, Crelle's Journal, vol. 150 (1920), pp. 79–106; p. 105. The present method applies also in the case of harmonic functions and harmonic polynomials; see Walsh, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 499–544; p. 513.

on $C_R$, then (6.4) can hold for $z$ on $C$ and with $R_1 > R$ for no sequence of polynomials $\{\pi_n(z)\}$.

When the function $f(z)$ is analytic on $C$, it is immaterial whether we consider the Tchebycheff polynomial for approximation to $f(z)$ on $C$ or merely *on the boundary of* $C$. For the functions whose absolute values appear in (6.1) and (6.2) take on their greatest moduli on the boundary. It is natural, then, to restrict ourselves to point sets $C$ which are boundaries of regions, and this is forced upon us if we consider, as we shall now do, functions with singularities interior to those regions.

THEOREM 6. *Let $r(z)$ be a rational function of $z$ with all of its singularities interior to $C$: $|z| = 1$, and let $\pi_n(z)$ be the Tchebycheff polynomial of degree $n$ for approximation to $r(z)$ on $C$. Let the function $f(z)$ (of Theorem 3) be analytic for $|z| < R > 1$ but have a pole on the circle $|z| = R$. Then the sequence $\{\pi_n(z)\}$ converges to the limit $f(z)$ for $|z| < R$, uniformly for $|z| < R_1 < R$.*

It is to be noted that $f(z)$ is analytic for $|z| = 1$, having on that circle the modulus $M$, and hence is analytic for $|z|$ less than some $R$ which is greater than unity. The Tchebycheff polynomials $\pi_n'(z)$ for approximation to $f(z)$ on $C$ satisfy the inequality

$$|f(z) - \pi_n'(z)| \leqq \frac{K}{R_1^n}, \quad z \text{ on } C,$$

by Theorem 5. Thus we can write

$$|r(z) - \pi_n'(z)| \leqq |r(z) - f(z)| + |f(z) - \pi_n'(z)| \leqq M + \frac{K}{R_1^n},$$

for $z$ on $C$. This inequality, holding for the polynomial $\pi_n'(z)$, yields us the inequality

$$|r(z) - \pi_n(z)| \leqq M + \frac{K}{R_1^n}, \quad z \text{ on } C,$$

which by (5.13) for $\epsilon_n = K/R_1^n$ yields

$$|f(z) - \pi_n(z)| \leqq \frac{K'}{R_1^n}$$

for $z$ on the circle $C'$: $|z| = \rho$, where $\rho$ is less than unity but is otherwise arbitrary. We apply Theorem 4 to the sequence $\{\pi_n(z)\}$ and to this circle $C'$, which yields the convergence of the sequence $\{\pi_n(z)\}$ for $|z| < \rho R_1$, uniformly for $|z| < \rho' < \rho R_1$. The numbers $\rho$ and $R_1$ are arbitrary, subject merely to the restriction of being respectively less than 1 and $R$, so Theorem 6 follows in its entirety.

Theorem 6 is obviously related to Theorem 3, but gives a more specific result concerning a more specific sequence of functions. There is an analogous

result related to Theorem 2, which involves the notion of *Tchebycheff polynomials with auxiliary conditions.*

We consider as before a function $\phi(z)$ continuous on a closed limited point set $C$. We consider all polynomials $p_n(z)$ of degree $n$ which take on prescribed values at $m$ points $P$ not belonging to $C$; such polynomials surely exist if $m$ is not greater than $n+1$. If the number of points of $C$ is sufficiently large (see §10) it can be shown that there exists a unique polynomial $\pi_n(z)$ belonging to this set of polynomials such that (6.1) is less than the expression (6.2) for any other polynomial $p_n(z)$ of the set. We shall call $\pi_n(z)$ the Tchebycheff polynomial for approximation to $\phi(z)$ on $C$ with the prescribed auxiliary conditions. It is permitted to consider the points $P$ not to be all distinct, that is, the auxiliary conditions may involve not merely the values of the polynomials $p_n(z)$ and $\pi_n(z)$ at the given points $P$, but also the values of the polynomials and their first $k_P$ derivatives at the points $P$.

The situation of Theorem 2 corresponds to approximation on the circle $C$ to the function $\phi(z) \equiv 0$, with the prescribed auxiliary conditions.

THEOREM 7. *Let the function $F(z)$ of Theorem 1 be analytic for $|z|<R>1$, but have a pole for which $|z|=R$. Then the sequence of Tchebycheff polynomials $\pi_n(z)$ (for approximation on $C$ to the function zero with the prescribed auxiliary conditions) converges to the function $F(z)$ for $|z|<R$, uniformly for $|z|<R_1<R$.*

There exist, by Theorem 4, polynomials $p_n(z)$ of respective degrees $n=0, 1, \cdots$, such that we have

$$|F(z) - p_n(z)| \leqq \frac{K}{R_1^n}, \quad z \text{ on } C,$$

where $R_1<R$ is arbitrary. For $n$ sufficiently large there exist polynomials $q_n(z)$ of respective degrees $n$, equal in the prescribed points $P$ to the functions $F(z)-p_n(z)$, and such that*

$$|q_n(z)| \leqq \frac{K'}{R_1^n}, \quad z \text{ on } C.$$

* See Walsh, these Transactions, vol. 30 (1928), pp. 307–332; Lemma, p. 319. The proof is there given only for the case $k_P=0$, but the general case can be treated in the same way. This same proof of (6.6) holds also for approximation to the real function $F(x)$ of the real variable $x$ on an interval $a \leqq x \leqq b$, and shows that the introduction of auxiliary conditions, requiring that the approximating polynomials $\pi_n(x)$ take on values of the function to be approximated at a finite number of preassigned points of the interval $(a, b)$, does not alter the degree of convergence of the sequence of Tchebycheff polynomials, if that is of the form corresponding to an inequality

$$|F(x) - \pi_n(x)| \leqq M \cdot \epsilon(n),$$

where $M$ is independent of $n$ and $\epsilon(n)$ approaches zero with $1/n$. Compare Jackson, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 259–262, who proves the corresponding fact for approximation to continuous functions by *trigonometric* polynomials with auxiliary conditions. If it is a question merely of establishing (6.6) in the present case, the treatment given in §12.2 of the present paper suffices.

Hence we have for $n$ sufficiently large

$$|F(z) - [p_n(z) + q_n(z)]| \leqq \frac{K + K'}{R_1^n}, \quad z \text{ on } C; \tag{6.5}$$

at each point $P$ we have

$$F(z) = p_n(z) + q_n(z),$$

so the polynomial $p_n(z)+q_n(z)$, like the function $F(z)$, satisfies the prescribed auxiliary conditions. Thus we have for the Tchebycheff polynomials

$$|F(z) - \pi_n(z)| \leqq \frac{K + K'}{R_1^n}, \quad z \text{ on } C, \tag{6.6}$$

and Theorem 7 follows from Theorem 4.

The notion of Tchebycheff polynomial is of course far more general than we have considered here. Let $C$ be any point set, and $F(z)$ a function defined on $C$, to be approximated on $C$ by linear combinations of functions of the given set $\{\chi_n(z)\}$. The function of the form

$$\Sigma_n(z) = a_1\chi_1(z) + \cdots + a_n\chi_n(z) \tag{6.7}$$

such that

$$\overline{\text{bound}}\,|F(z) - \Sigma_n(z)|, \quad z \text{ on } C, \tag{6.8}$$

is least is called the Tchebycheff polynomial $\Sigma_n'(z)$ of order $n$ for approximation to $F(z)$ on $C$. It is also possible not to admit all functions of form (6.7) into consideration, but to require for instance auxiliary conditions as in the present paragraph.

The Tchebycheff polynomial $\Sigma_n(z)$ may or may not exist, and if it exists may or may not be unique.* It is always true, however, that the sequence

$$M_n = \overline{\text{bound}}\,|F(z) - \Sigma_n'(z)|, \quad z \text{ on } C,$$

decreases monotonically and hence approaches a limit, for the function $\Sigma_n'(z)$ may also be considered a function $\Sigma_{n+1}(z)$. It is also true that if the lower limit of (6.8) for all admissible functions $\Sigma_n(z)$, $n=1, 2, \cdots$, is a number $M$, then we have

$$\lim_{n\to\infty} M_n = M.$$

For $\lim_{n\to\infty} M_n$ can be greater than no expression (6.8).

**7. Regions more general than circles; approximation with auxiliary conditions.** Theorems 1-3 hold with only obvious changes for simply connected regions† other than circles. The new theorems are proved from the old ones by means of conformal mapping of the new regions on the circle.

* Compare Haar, Mathematische Annalen, vol. 78 (1918), pp. 294–311.

† A region is an open connected point set.

THEOREM 8. *Let $B$ be an arbitrary simply connected region of the $w$-plane with the boundary $C$. Among all functions analytic interior to $B$ each of which at $k$ given points $P$ of $B$ takes on preassigned values for itself and its first $k_P$ derivatives, there exists a unique function $G(w)$ which has the smallest least upper bound for its modulus in $B$. The modulus of this function $G(w)$ approaches a constant value $M$ as $w$ remaining in $B$ approaches $C$.*

*If the functions $G_n(w)$ analytic in $B$ satisfy the auxiliary conditions at the $k$ points $P$, then*

$$\lim_{n\to\infty} [\overline{\text{bound}}\, G_n(w),\ w \text{ in } B] = M$$

*implies $\lim_{n\to\infty} G_n(w) = G(w)$ for $w$ in $B$, uniformly for $w$ on any closed point set interior to $B$.*

Theorem 8 follows directly from Theorems 1 and 2, for under conformal mapping of $B$ onto the unit circle in the $z$-plane, a function which takes on prescribed values at definite points of $B$ corresponds to a function which takes on prescribed values at definite points in the interior of the unit circle, and conversely. Prescription of a certain number of derivatives at a point in the $w$-plane implies prescription of the same number of derivatives at the corresponding point of the $z$-plane, and conversely. Moreover, if $w$ in $B$ approaches $C$, then $z$, lying in the unit circle, approaches the circumference, and conversely.

A degenerate case occurs here if the region $B$ is the entire $w$-plane, or the entire plane except for a single point. In this case the conformal map is impossible. If there exists a constant satisfying the prescribed auxiliary conditions, Theorem 8 is valid as stated, $M$ being the absolute value of this constant. If no such constant exists, there is no function which satisfies the auxiliary conditions whose modulus in $B$ has a finite upper bound.

If the region $B$ is bounded by an analytic Jordan curve $C$, a more explicit result than Theorem 8 can be established for the Tchebycheff polynomials with the prescribed auxiliary conditions; this new result is a generalization of Theorem 7.

THEOREM 9. *If in Theorem 8 the region $B$ is the interior of an analytic Jordan curve $C$, then there exists some $C_R$ such that the function $G(w)$ is analytic for $w$ interior to $C_R$ but has a singularity on $C_R$. Then the sequence of Tchebycheff polynomials $\{\tau_n(w)\}$ (for approximation on $C$ to the function zero with the prescribed auxiliary conditions) converges to the function $G(w)$ for $w$ interior to $C_R$, uniformly for $w$ on an arbitrary closed point set interior to $C_R$.*

The proof here follows directly the proof of Theorem 7 and is omitted.

A less explicit result can be obtained for much more general regions.

THEOREM 10. *If in Theorem* 8 *the finite region B is such that its boundary is also the boundary of an infinite region, then the sequence* $\{\tau_n(w)\}$ *of Tchebycheff polynomials for approximation to the function zero on C with the given auxiliary conditions converges to the function* $G(w)$ *in B, uniformly on any closed point set interior to B.*

Let $B_\mu$ be a sequence of closed regions bounded by analytic Jordan curves $C_\mu$ converging monotonically to $C$. That is, $B$ and $C_{\mu+1}$ are interior to $C_\mu$, and no point not belonging to the closed region $B$ lies interior to all the $C_\mu$ unless such a point cannot be joined to the point at infinity by a broken line which does not meet $C$. Let $\tau_{\mu n}(w)$ be the Tchebycheff polynomial of degree $n$ for the region $B_\mu$, that is, having the least maximum modulus in that closed region and satisfying the given auxiliary conditions; these auxiliary conditions are simply that the polynomial should take on certain values at the given points $P$ interior to $B$ and do not depend on $\mu$. Let $M_{\mu n}$ denote the maximum modulus of $\tau_{\mu n}(w)$ in $B_\mu$—that is, on $C_\mu$.

We have the relation

$$M_{\mu+1,n} \leqq M_{\mu n}, \tag{7.1}$$

for $\tau_{\mu n}(w)$ is a polynomial of degree $n$ which satisfies the given auxiliary conditions, and its maximum modulus on $C_{\mu+1}$ is less than or equal to $M_{\mu n}$, its maximum modulus on $C_\mu$, since $C_{\mu+1}$ is interior to $C_\mu$; the equality can occur only if $\tau_{\mu n}(w)$ is a constant. Hence the maximum modulus of $\tau_{\mu+1,n}(w)$ on $C_{\mu+1}$ is less than $M_{\mu n}$, or in case $\tau_{\mu+1,n}(w)$ is the same as $\tau_{\mu n}(w)$ and a constant, is equal to $M_{\mu n}$. We have also the relation

$$M_{\mu,n+1} \leqq M_{\mu n}, \tag{7.2}$$

for $\tau_{\mu n}(w)$ can be considered a polynomial of degree $n+1$, and satisfies the auxiliary conditions.

The region $B$ is the kernel of the regions $B_\mu$, and it follows* from the properties of the conformal map and from §3 that as $\mu$ becomes infinite the functions $G_\mu(w)$ (of Theorem 8) corresponding to the regions $B_\mu$ approach the function $G(w)$ throughout $B$, uniformly on any closed point set interior to $B$. For map both $B$ and $B_\mu$ onto the unit circle in the $z$-plane so that a definite point and direction in the $w$-plane correspond to one and the same point and direction in the $z$-plane. The difference between the mapping functions is small, uniformly in an arbitrary closed region interior to $B$, and can be

* Carathéodory, Mathematische Annalen, vol. 72 (1912), pp. 107–144, chapter III.

made as small as desired by choosing $\mu$ sufficiently large,* and the same is true of the derivatives of the mapping functions. The auxiliary conditions for the Tchebycheff polynomials thus differ only slightly in the two cases, approximation in $B$ and in $B_\mu$, when we interpret the problem in the $z$-plane after mapping these respective regions on the unit circle. We now obtain the functions $G(w)$ and $G_\mu(w)$ by mapping the unit circle on $B$ and $B_\mu$ respectively, and these mapping functions differ but little from each other on an arbitrary closed point set interior to the unit circle; the difference can be made uniformly and arbitrarily small by choosing $\mu$ sufficiently large. Hence by §3 the difference between the functions $G(w)$ and $G_\mu(w)$ can be made as small as desired on an arbitrary closed point set interior to $B$, by choosing $\mu$ sufficiently large. It follows likewise from §3 that

$$\lim_{\mu\to\infty} M_\mu = M, \tag{7.3}$$

where $M$ and $M_\mu$ are respectively the maximum moduli of $G(w)$ and $G_\mu(w)$ in $B$ and in $B_\mu$.

By virtue of the two relations (7.1) and (7.2), it follows that $\lim_{\mu,n\to\infty} M_{\mu n}$ exists, and that this limit is equal to

$$\lim_{\mu\to\infty} [\lim_{n\to\infty} M_{\mu n}]. \tag{7.4}$$

By the method of proof of Theorem 9 we have $\lim_{n\to\infty} M_{\mu n} = M_\mu$, so by (7.3) we have (7.4) equal to $M$. Hence, by the monotonic character of the sequence $M_{\mu n}$, we have also

$$\lim_{n\to\infty} [\lim_{\mu\to\infty} M_{\mu n}] = M. \tag{7.5}$$

---

* The condition that the boundary of $B$ should also be the boundary of an infinite region is precisely the form of condition necessary for the application of Carathéodory's theorem, when we approximate $B$ externally, as here with the regions $B_\mu$. The region $B$ may still be fairly complicated and in particular its boundary may separate the plane into more than two regions. The region $B$ may be, for example, a strip which is closed at one end and approaches a circle by winding about it externally infinitely often. It is because the boundary of $B$ may separate the plane into more than two regions and because of the corresponding difficulties after conformal mapping that the function $f(z)$ of Theorem A is so restricted as to have no singularities except in $B$ and in the infinite region of which $C$ is the boundary; compare Theorem 15 below.

Compare Julia, loc. cit., chapter III.

The Tchebycheff polynomial with auxiliary conditions for approximation on an arbitrary closed limited point set $C$, to a function $F(z)$ continuous on $C$ and analytic in the interior points of $C$, is precisely the same as the Tchebycheff polynomial with those auxiliary conditions for approximation on the boundary $C'$ of the infinite region $B'$ consisting of all points which can be joined with the point at infinity by broken lines not meeting $C$. For the maximum on $C$ of such a function as $|F(z)-\pi_n(z)|$, where $\pi_n(z)$ is the Tchebycheff polynomial, occurs on $C'$. Special cases of this fact are indicated by Julia for his particular auxiliary conditions.

The quantity $M_n'$ here in square brackets is equal to the maximum modulus $M_n''$ on $C$ of $\tau_n(w)$. For if we had $M_n'' < M_n'$, we should be able to find some region $B_\mu$ for which the Tchebycheff polynomial $\tau_{\mu n}(w) = \tau_n(w)$ likewise had in $B_\mu$ a maximum modulus less than $M_n'$, in contradiction with (7.1). The relation $M_n' < M_n''$ is likewise impossible, for the maximum modulus of $\tau_{\mu n}(w)$ in $B$ is less than or equal to $M_{\mu n}$, for $B$ is interior to $C_\mu$, hence $M_n''$ is less than or equal to $M_{\mu n}$.

The relation (7.5) can be written

$$\lim_{n\to\infty} M_n'' = M,$$

which by virtue of Theorem 8 gives now Theorem 10. We might well state a more general result, for we have

$$\lim_{\mu,n\to\infty} M_{\mu n} = M,$$

so that we have proved the existence of the double limit

$$\lim_{\mu,n\to\infty} \tau_{\mu n}(w) = G(w)$$

throughout $B$, uniformly on any closed point set interior to $B$.

The most interesting auxiliary conditions in connection with Theorems 8–10 are

$$\tau_n(0) = 0, \quad \tau_n'(0) = 1, \tag{7.6}$$

where the origin $w=0$ is interior to $C$. Map the region $B$ onto the interior of the unit circle $|z| \leqq 1$ by means of the function $w=\phi(z)$, $z=\psi(w)$, so that $w=0$ corresponds to $z=0$; the auxiliary conditions are equivalent to $\tau_n[\phi(0)]=0$, $\tau_n'[\phi(0)]=1$. This second relation is, by the formula

$$\frac{d\tau_n}{dw} = \frac{d\tau_n}{dz}\frac{dz}{dw},$$

equivalent to

$$\frac{d\tau_n}{dz} = a, \text{ some constant not zero.}$$

The problem of approximation in $B$ is equivalent to the problem of approximation in the unit circle of the $z$-plane by functions of the form $F_n(z) = az + a_2z^2 + a_3z^3 + \cdots$, where $a$ is prescribed but the values $a_i$ are not prescribed; the modulus of $F_n(z)$ for $|z|<1$ is not greater than

$$\overline{\text{bound}}_{|z|<1} |a + a_2z + a_3z^2 + \cdots|$$

whose minimum value (compare Carathéodory's Lemma) is of course $|a|$.

The corresponding function $F(z)$ is $az$, and the function $G(w)$ is $az=a\psi(w)$, which maps $B$ onto a circle in the $w$-plane.

Our general problem of Theorems 8–10 for the particular auxiliary conditions (7.6) has been studied in detail by Julia (loc. cit.), who has obtained for that case the principal results, with the exception of Theorem 9, that we have obtained in the more general case.

**8. Regions more general than circles; approximation to a rational function.** Just as we have extended Theorems 1 and 2 to the case of regions more general than circles, we shall now similarly extend Theorem 3:

THEOREM 11. *Let $B$ be an arbitrary simply connected region of the $w$-plane with the boundary $C$, and let $s(w)$ be a rational function of $w$ with all of its singularities interior to $B$. Of all functions $g(w)$ analytic interior to $B$, there exists a unique function $g(w)$ such that*

$$\overline{\text{bound}}\,|\,s(w)-g(w)\,|\ \textit{for } w \textit{ in } B \textit{ approaching } C \tag{8.1}$$

*is least; the function $s(w)-g(w)$ is of constant modulus $M$ on $C$,—that is, the modulus of $s(w)-g(w)$ approaches a constant value as $w$ in $B$ approaches $C$.*

*If the functions $g_n(w)$ are analytic for $w$ in $B$, then*

$$\lim_{n\to\infty}\Big[\overline{\underset{\substack{w\to C\\ w \text{ in } B}}{\text{bound}}}\,|\,s(w)-g_n(w)\,|\,\Big]=M \tag{8.2}$$

*implies*

$$\lim_{n\to\infty} g_n(w)=g(w)$$

*for $w$ in $B$, uniformly for $w$ on any closed point set in $B$.*

This theorem follows without difficulty from Theorem 3 by conformal mapping of the interior of $B$ onto the interior of the unit circle in the $z$-plane. Under the map the rational function $s(w)$ corresponds to a function of $z$:

$$s(w)=r(z)-s_1(z), \tag{8.3}$$

where $r(z)$ is a rational function of $z$ and $s_1(z)$ is analytic for $|z|<1$. There exists a function $f(z)$ for which

$$\overline{\underset{\substack{|z|\to 1\\ |z|<1}}{\text{bound}}}\,|\,r(z)-f(z)\,|$$

is least. Under the conformal map, a function analytic interior to $B$ corresponds to a function analytic for $|z|<1$ and conversely, and $|z|\to 1$ implies $w\to C$ and conversely, so this upper bound is the same as (8.1), if we set

$$s_1(z)+g(w)=f(z). \tag{8.4}$$

A unique minimizing function $f(z)$ leads to a unique minimizing function $g(w)$ and conversely. We can similarly identify (4.4) and (8.2) by setting

$$s_1(z) + g_n(w) = f_n(z);$$

equation (8.2) implies $\lim_{n\to\infty} f_n(z) = f(z)$, which implies $\lim_{n\to\infty} g_n(w) = g(w)$, uniformly as stated, so Theorem 11 is completely established.

A degenerate case of Theorem 11 deserves special mention, when $B$ is the entire $w$-plane or the entire plane except for a single point $w=\alpha$. The first case has no meaning, for (8.1) of necessity refers to the boundary $C$ and here $C$ is non-existent. The second case is not difficult if we interpret (8.1) literally. All the functions $g(w)$ which concern us are constants, and the least (8.1) is given by setting $g(w) \equiv s(\alpha)$, so that (8.1) is zero. The second part of the theorem is also exact. In the future we shall not mention the possibility of the degeneracy of the region $B$; the reader can easily treat the new situation, for instance in Theorem 14.

In a special case we have a result more specific than Theorem 11:

THEOREM 12. *If in Theorem* 11 *the region $B$ is the interior of an analytic Jordan curve $C$, the function $g(w)$ is analytic for $w$ interior to some $C_R$, but has a singularity on $C_R$. Then the sequence of Tchebycheff polynomials $\omega_n(w)$ for approximation on $C$ to the function $s(w)$ converges to the function $g(w)$ for $w$ interior to $C_R$, uniformly for $w$ on an arbitrary closed point set interior to $C_R$.*

In the notation just used, we have by Theorem 4 the existence of a set of polynomials $\omega_n'(w)$ of respective degrees $n$ such that

$$|\, g(w) - \omega_n'(w) \,| \leqq \frac{M'}{R_1^n}, \text{ for } w \text{ on } B + C,$$

where $R_1$ is an arbitrary positive number less than $R$. The equation

$$|\, r(z) - f(z) \,| = M, \; w \text{ on } C,$$

implies then

$$|\, s(w) - \omega_n'(w) \,| \leqq M + \frac{M'}{R_1^n},$$

for $s(w) = r(z) - f(z) + g(w)$. We may obviously write, then, by the definition of the Tchebycheff polynomials $\omega_n(w)$,

$$|\, s(w) - \omega_n(w) \,| \leqq M + \frac{M'}{R_1^n};$$

we rewrite this inequality in the form

$$|\, r(z) - [f(z) - g(w) + \omega_n(w)] \,| \leqq M + \frac{M'}{R_1^n},$$

which holds for $w$ in $B$ approaching $C$, and hence for $|z|<1$, $|z|\to 1$. This inequality implies, by (5.13),

$$|[f(z)-g(w)+\omega_n(w)]-f(z)| \leqq \frac{M''}{R_1^n},$$

for $|z|=\rho<1$, which is the same as

$$|g(w)-\omega_n(w)| \leqq \frac{M''}{R_1^n}$$

for $w$ on an arbitrary Jordan curve $C'$ interior to $C$, where of course $M''$ depends on $C'$. The fact that this inequality holds for an arbitrary $R_1<R$ and an arbitrary curve $C'$ interior to $C$ yields, by virtue of Theorem 4, Theorem 12 as stated. For when the curve $C'$ approaches the curve $C$ uniformly, then the curve $C'_{R_1}$ (image in the $w$-plane of the circle $|z|=R_1>1$ if the exterior of $C'$ is mapped onto the exterior of $|z|=1$ so that the points at infinity correspond to each other) approaches the curve $C_{R_1}$ uniformly. This follows from Carathéodory's results on conformal mapping.

The analogue of Theorem 10 will now be proved.

THEOREM 13. *If in Theorem* 11 *the finite region $B$ is such that its boundary $C$ is also the boundary of an infinite region, then the sequence $\{\omega_n(w)\}$ of Tchebycheff polynomials for approximation on $C$ to the function $s(w)$ converges throughout $B$ to the function $g(w)$, uniformly on any closed point set interior to $B$.*

We consider the sequence of analytic Jordan curves $C_\mu$ exterior to $B$ used in the proof of Theorem 10. The problem of approximation to $s(w)$ on $C_\mu$ will be studied in connection with the problem of approximation to $s(w)$ on $C$. Let $g^{(\mu)}(w)$ denote the function $g(w)$ of Theorem 11 which corresponds to the former problem, and $r^{(\mu)}(z)$ the rational function of $z$ which gives the equivalent problem of approximation in the unit circle in the $z$-plane:

$$(8.5)\qquad s(w)=r^{(\mu)}(z)-s_1^{(\mu)}(z), \qquad s(w)-g^{(\mu)}(w)=r^{(\mu)}(z)-f^{(\mu)}(z),$$

where $w=\phi_\mu(z)$, the function which maps $B_\mu$ onto $|z|<1$ and has auxiliary conditions independent of $\mu$.

The singularities of the function

$$(8.6)\qquad r(z)-f(z)=s(w)-g(w)$$

(where $w=\phi(z)$, the function which maps $B$ onto $|z|<1$ and has the same auxiliary conditions as the $\phi_\mu(z)$) for $|z|<1$ differ but slightly from those of the function $r^{(\mu)}(z)-f^{(\mu)}(z)$. For these singularities depend merely on the singularities of $s(w)$ and on the values of the mapping functions and their derivatives in the neighborhoods of those singularities. The singularities vary continuously with the mapping functions (that is, the coefficients of

the Laurent series for development about the singularities vary continuously) and so does $r^{(\mu)}(z)-f^{(\mu)}(z)$, by §3 and equation (4.6). In fact we have

$$\lim_{\mu\to\infty}\left[r^{(\mu)}(z)-f^{(\mu)}(z)\right]=r(z)-f(z)$$

uniformly for $|z|\leqq 1$, except in the neighborhoods of the singularities of $r(z)$, for it is true that $\lim_{\mu\to\infty}\phi_\mu(z)=\phi(z)$ uniformly for $|z|\leqq\rho<1$.

Let an arbitrary positive $\epsilon$ be given. Choose $\mu$ so large that we have

$$M^{(\mu)}<M+\frac{\epsilon}{3},$$

where $M^{(\mu)}=|r^{(\mu)}(z)-f^{(\mu)}(z)|$ for $|z|=1$. Choose $\mu$ so large that we have also

$$|s(w)-g^{(\mu)}(w)|<M^{(\mu)}+\frac{\epsilon}{3}$$

for $w$ on $C$ and hence for $w$ in some neighborhood of $C$. This choice is possible, for we have as $\mu$ becomes infinite

$$r^{(\mu)}(z)-f^{(\mu)}(z)=s(w)-g^{(\mu)}(w)\to r(z)-f(z)$$

uniformly in the neighborhood of the circle $|z|=1$ (indeed throughout $|z|\leqq 1$ except in the neighborhood of the singularities of $r(z)$), where the second member of the equality is transformed by $w=\phi_\mu(z)$. Under the transformation, it is to be noticed, the point set $C$ corresponds to a closed point set in the neighborhood of $|z|=1$, and this closed point set approaches $|z|=1$ uniformly when $\mu$ becomes infinite.

It is now possible, by Theorem 12, to choose the polynomial $\tau(w)$ so that for a particular $\mu$ (but sufficiently large to ensure the two inequalities just written) we have

$$|g^{(\mu)}(w)-\tau(w)|<\frac{\epsilon}{3}$$

for $w$ in some neighborhood of $C$; in fact this inequality can be satisfied on and within $C_\mu$. That is to say, we have

$$|s(w)-\tau(w)|<M+\epsilon \tag{8.7}$$

for $w$ in the neighborhood of $C$.

Inequality (8.7) informs us that there exists some polynomial $\tau(w)$ such that the left-hand member is less than $M+\epsilon$. It follows that the monotonically decreasing sequence

$$\max|s(w)-\omega_n(w)|,\quad \text{for } w \text{ on } C,$$

which approaches a limit, can approach no limit greater than $M$. This sequence can obviously, by Theorem 11, approach no limit less than $M$. Hence the limit of the sequence is $M$, and Theorem 13 is a consequence of Theorem 11.

The general problem of Theorem 11 is especially interesting if $s(w)$ is of the form $a/w$, where $a$ is a constant not zero. Under the transformation $w=\phi(z)$ we find that the function $r(z)$ of (8.3) is of the form $b/z$, if $\phi(0)=0$, so we refer to the problem of Theorem 3 where $r(z)=b/z$, $b\neq 0$. In this case we have $f(z)\equiv 0$, so

$$s(w) - g(w) = r(z) - f(z) = \frac{b}{z}$$

is a function which maps $B$ onto the exterior of a circle of radius $|b|$ in the $z$-plane, in such a manner that the point $w=0$ corresponds to the point $z=\infty$.

**9. Approximation to more general functions.** Certain generalizations of the results previously obtained are entirely obvious. Let the function $r_1(z)$ have no singularities interior to $C$: $|z|=1$ other than a finite number of poles,

$$r_1(z) = r(z) + r_2(z),$$

where $r(z)$ is a rational function of $z$ with all of its singularities interior to $C$, and where $r_2(z)$ is analytic interior to $C$. Then the reader can easily prove (compare Theorem 3): *Among all functions $f'(z)$ analytic for $|z|<1$, there exists a unique function such that*

$$\overline{\operatorname{bound}}_{\substack{|z|\to 1\\ |z|<1}} |\, r_1(z) - f'(z)\,|$$

*is least; this function $f'(z)$ can be written $f(z)+r_2(z)$, where $f(z)$ is rational. The function $r_1(z)-f'(z)$ is of constant modulus $M$ on $C$. If the functions $f_n'(z)$ are analytic for $|z|<1$, then*

$$\lim_{n\to\infty} \left[\overline{\operatorname{bound}}_{\substack{|z|\to 1\\ |z|<1}} |\, r_1(z) - f_n'(z)\,|\,\right] = M$$

*implies* $\lim_{n\to\infty} f_n'(z)=f'(z)$ *for* $|z|<1$, *uniformly for* $|z|\leqq\rho<1$.

Later theorems of §8 likewise have analogues in this new situation; we sketch the modifications to be made. We assume (compare Theorem 11) $s_1(w)$ to have no singularities in $B$ other than a finite number of poles:

$$s_1(w) = s(w) + s_2(w),$$

where $s(w)$ is a rational function of $w$ with all of its singularities in $B$ and $s_2(w)$ is analytic in $B$. *Of all functions $g'(w)$ analytic interior to $B$, there exists a unique function such that*

$$\overline{\operatorname{bound}}_{\substack{w\to C\\ w \text{ in } B}} |\, s_1(w) - g'(w)\,|$$

*is least; the function $s_1(w)-g'(w)$ is of constant modulus $M$ on $C$—that is, the modulus of $s_1(w)-g'(w)$ approaches a constant value $M$ as $w$ in $B$ approaches $C$. If the functions $g_n'(w)$ are analytic for $w$ in $B$, then*

$$\lim_{n\to\infty} \left[\overline{\text{bound}}_{\substack{w\to C\\ w \text{ in } B}} \left| s_1(w) - g_n'(w) \right|\right] = M$$

*implies $\lim_{n\to\infty} g_n'(w)=g'(w)$ for $w$ in $B$, uniformly for $w$ on any closed point set in $B$.*

Theorem 12 holds unchanged if $s(w)$ and $g(w)$ are replaced by $s'(w)$ and $g'(w)$ respectively. Theorem 13 holds also if, in our present notation, the function $s_2(w)$ can be uniformly approximated in $B+C$ as closely as desired by a polynomial in $w$, and thus holds in particular if $B$ is a Jordan region and $s_2(w)$ is analytic in $B$, continuous in $B+C$.

The situation of Theorem 1 is also easy to generalize. Let the function $\Phi(z)$ be given analytic interior to $C$: $|z|<1$; we consider the problem of determining the function $\phi(z)$ also analytic for $|z|<1$, which takes on prescribed values at preassigned points $\alpha_1, \alpha_2, \cdots, \alpha_k$ interior to $C$, and such that

$$\overline{\text{bound}} \left| \Phi(z) - \phi(z) \right|, \text{ for } |z| < 1, \tag{9.1}$$

is least. Any function $F(z)$ analytic for $|z|<1$ which takes on the values $\Phi(\alpha_i)-\phi(\alpha_i)$ at the points $\alpha_i$ can be written in the form

$$F(z) = \Phi(z) - \phi(z), \tag{9.2}$$

where this equation defines $\phi(z)$. Conversely, if $\phi(z)$ is given, a corresponding function $F(z)$ analytic for $|z|<1$ and taking on the values $\Phi(\alpha_i)-\phi(\alpha_i)$ at the points $\alpha_i$ can be considered defined by (9.2). The entire problem of the existence, uniqueness, and determination of $\phi(z)$ is equivalent to the problem of the existence, uniqueness, and determination of the function $F(z)$ of Theorem 1 and can then be considered solved. The reader should have no difficulty in writing down the analogues of Theorems 1, 2, 7, 8, 9, 10.

Suppose the function $\Phi(z)$ is an arbitrary function of $z$ with not more than a finite number of poles and no other singularities interior to $C$: $|z|=1$; we consider as in Theorem B the problem of determining the function $\phi(z)$ analytic interior to $C$ which takes on prescribed values at preassigned points $\alpha_1, \alpha_2, \cdots, \alpha_k$ interior to $C$ and is such that

$$\overline{\text{bound}}_{\substack{|z|\to 1\\ |z|<1}} \left| \Phi(z) - \phi(z) \right| \tag{9.3}$$

is least. Let $\beta_1, \beta_2, \cdots, \beta_l$ denote the poles of $\Phi(z)$ interior to $C$, each pole

enumerated the number of times corresponding to its multiplicity; then (9.3) is precisely the same as

$$(9.4)\qquad \overline{\underset{\substack{|z|\to 1\\ |z|<1}}{\text{bound}}} \left| [\Phi(z)-\phi(z)]\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_l)}{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_l z)}\right|.$$

This last expression shows that our problem is equivalent to the problem of approximating the function

$$\Phi_1(z)=\Phi(z)\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_l)}{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_l z)},$$

which is analytic interior to $C$, by the function

$$\phi_1(z)=\phi(z)\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_l)}{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_l z)},$$

also analytic interior to $C$, a problem of the type just considered. Determination of the function $\phi(z)$ with the prescribed auxiliary conditions leads to the determination of $\phi_1(z)$ with the conditions

$$\phi_1(\alpha_i)=\phi(\alpha_i)\frac{(\alpha_i-\beta_1)(\alpha_i-\beta_2)\cdots(\alpha_i-\beta_l)}{(1-\bar\beta_1\alpha_i)(1-\bar\beta_2\alpha_i)\cdots(1-\bar\beta_l\alpha_i)} \qquad (i=1,2,\cdots,k),$$

$$\phi_1(\beta_i)=0 \qquad (i=1,2,\cdots,l,\quad \beta_i\neq\alpha_j);$$

and reciprocally, determination of $\phi_1(z)$ with these auxiliary conditions leads to the determination of $\phi(z)$ satisfying the prescribed conditions. Obvious alterations in these conditions are here necessary if the $\beta_i$ are not all distinct, or if derivatives of $\phi(z)$ are prescribed as well as functional values. The usual conventions suffice, however, even if some of the $\alpha_i$ coincide with some of the $\beta_i$.

We add a remark on the degrees of the functions involved. The function $F(z)=\Phi_1(z)-\phi_1(z)$ is the minimizing function (as in Theorem 1) corresponding to $k+l$ auxiliary conditions and hence is of degree less than $k+l$. We have

$$\Phi(z)-\phi(z)=\frac{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_l z)}{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_l)}F(z),$$

which is therefore of degree less than $k+2l$.

We state the result as a theorem not in the $z$-plane but after conformal mapping onto the $w$-plane; the detailed reasoning involved is already familiar to the reader.

THEOREM 14. *Let $B$ be an arbitrary simply connected region of the $w$-plane, with the boundary $C$. Let $\Psi(w)$ be analytic in $B$ except possibly for a finite number of poles. Let $\psi(w)$ denote generically a function analytic in $B$ which at*

*$k$ given points $P$ of $B$ has preassigned values for itself and its first $k_P$ derivatives. Then a unique function $\psi(w)$ exists so that*

$$\overline{\operatorname{bound}}_{\substack{w \to C \\ w \text{ in } B}} | \Psi(w) - \psi(w) |$$

*is least. The modulus of $\Psi(w)-\psi(w)$ approaches a constant value $M$ as $w$ remaining in $B$ approaches $C$.*

*If the functions $\psi_n(w)$ analytic in $B$ satisfy the auxiliary conditions, then*

$$\lim_{n \to \infty} \left[ \overline{\operatorname{bound}}_{\substack{w \to C \\ w \text{ in } B}} | \Psi(w) - \psi_n(w) | \right] = M$$

*implies*

$$\lim_{n \to \infty} \psi_n(w) = \psi(w)$$

*for $w$ in $B$, uniformly on any closed point set interior to $B$.*

Theorem 14 can be interpreted as a direct extension of Theorem 1, where not only finite but also infinite values of the function may be assigned at the $k'$ points $P$ (i.e., analogous to the points $P$ of Theorem 8) interior to $B$. These $k'$ points $P$ are of course the $k$ points $P$ of Theorem 14 plus the poles of $\Psi(w)$ interior to $B$. That is to say, we consider the class of functions $\Psi(w)-\psi(w)$ of the form

$$a_{-q}(w-p)^{-q} + a_{-q+1}(w-p)^{-q+1} + \cdots + a_0 + a_1(w-p) + \cdots + a_m(w-p)^m + b_{m+1}(w-p)^{m+1} + \cdots,$$

where at each of the $k'$ given points $P: w=p$, the numbers $a_i$ (depending on $P$) are prescribed and the numbers $b_i$ are not prescribed; the restriction $m>0$ or $m \geqq 0$ is not made. We naturally consider as a measure of the modulus of $F(z)=\Psi(w)-\psi(w)$ not the upper limit in $B$ of the modulus of the function, but as in Theorem 11,

$$\overline{\operatorname{bound}}_{\substack{w \to C \\ w \text{ in } B}} | \Psi(w) - \psi(w) | .$$

There is no difference in these measures of the modulus of $F(z)$ if $F(z)$ is analytic interior to $B$ (mapped on the unit circle in the $z$-plane) but of course there is a difference if the auxiliary conditions involve poles of $F(z)$. There exists a unique function $F(z)=\Psi(w)-\psi(w)$ of this class whose least upper bound for $w$ in $B$ approaching $C$ is less than the least upper bound for $w$ in $B$ approaching $C$ for any other function of the class. This function is rational in $z$ (that is, when $B$ is mapped on the unit circle in the $z$-plane) and of constant modulus $M$ on $C$, in the sense that

$$\lim_{\substack{w \to C \\ w \text{ in } B}} | \Psi(w) - \psi(w) | = M.$$

An arbitrary rational function $r(z)$ of constant modulus on $\gamma$: $|z|=1$ can be written in the form

$$r(z) = K\frac{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_\mu)(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_\nu z)}{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)\cdots(1-\bar{\alpha}_\mu z)(z-\beta_1)(z-\beta_2)\cdots(z-\beta_\nu)}.$$

Let $r(z)$ not be identically zero and let $\alpha_1, \alpha_2, \cdots, \alpha_\mu$ denote the zeros of $r(z)$ interior to $\gamma$ and $\beta_1, \beta_2, \cdots, \beta_\nu$ the poles of $r(z)$ interior to $\gamma$. Then the function

$$r(z)\frac{(1-\bar{\alpha}_1 z)\cdots(1-\bar{\alpha}_\mu z)(z-\beta_1)\cdots(z-\beta_\nu)}{(z-\alpha_1)\cdots(z-\alpha_\mu)(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_\nu z)}$$

is analytic and different from zero on and within $\gamma$ and of constant modulus on $\gamma$, hence is a constant.

It is perhaps worth while to give the details of the proof of the following addition to Theorem 14:

THEOREM 15. *Suppose $B$ is a finite region whose boundary $C$ is also the boundary of an infinite region; and suppose when we write*

$$\Psi(w) = \Psi_1(w) + \Psi_2(w),$$

*where $\Psi_1(w)$ is a rational function of $w$ and has in $B$ precisely the singularities of $\Psi(w)$ and is analytic on and exterior to $C$, the function $\Psi_2(w)$ can be uniformly approximated on $B+C$ as closely as desired by a polynomial in $w$; then the sequence $\{\omega_n(w)\}$ of Tchebycheff polynomials for approximation on $C$ to the function $\Psi(w)$ with arbitrary given auxiliary conditions, converges to the function $\psi(w)$ of Theorem 14 throughout $B$, uniformly on any closed point set interior to $B$.*

*If moreover $C$ is an analytic Jordan curve, and if the function $\psi(w)$ is analytic on and interior to $C_R$, then the sequence $\{\omega_n(w)\}$ converges uniformly on and within $C_R$.*

We prove first the second part of Theorem 15. We have by Theorem 4 the existence of a sequence of polynomials $\omega_n'(w)$ of respective degrees $n$ such that

$$|\psi(w) - \omega_n'(w)| \leqq \frac{M'}{R^n} \text{ for } w \text{ on } B+C.$$

These polynomials $\omega_n'(w)$ are chosen also to satisfy the given auxiliary conditions; the function $\psi(w)$ satisfies those conditions and hence (as in the proof of Theorem 7) the polynomials $\omega_n'(w)$ can likewise be chosen to satisfy them. We have also

$$\lim_{\substack{w \to C \\ w \text{ in } B}} |\Psi(w) - \psi(w)| = M,$$

and thus derive

$$|\Psi(w) - \omega_n'(w)| \leqq M + \frac{M'}{R^n}, \text{ for } w \text{ on } C.$$

It follows by the definition of the Tchebycheff polynomial $\omega_n(w)$ that we have

$$|\Psi(w) - \omega_n(w)| \leqq M + \frac{M'}{R^n}, \text{ for } w \text{ on } C;$$

the result as stated in the theorem now follows as in the proof of Theorem 12. It will be noticed that by (9.4) an inequality of type (5.13) holds indifferently for approximation to functions of type $\Phi(z)$ as used in (9.3) by a sequence of functions of type $\phi(z)$, or for approximation to functions of type $\Phi_1(z)$ by a sequence of functions of type $\phi_1(z)$.

We turn now to the proof of the first part of Theorem 15. Consider new auxiliary conditions, so that the prescribed values for the new polynomial $\omega''(w)$ now to be introduced differ from those of $\omega_n(w)$ by the values of $\Psi_2(w)$ in the respective points:

$$\omega''(w) + \Psi_2(w) = \omega_n(w), \text{ when } w \text{ is a point } P.$$

It follows from the proof of Theorem 13 by the use of the reasoning used in the proof of Theorem 7, that a polynomial $\omega''(w)$ of some degree $m$ exists which satisfies the new auxiliary conditions and is such that we have

$$(9.5) \qquad |[\Psi(w) - \Psi_2(w)] - \omega''(w)| \leqq M + \frac{\epsilon}{2}, \text{ for } w \text{ on } C,$$

where $\epsilon$ is arbitrary but preassigned. For it is obvious that if in Theorem 14 the function $\Psi(w)$ is replaced by $\Psi_1(w) = \Psi(w) - \Psi_2(w)$, and the auxiliary conditions satisfied by $\psi(w)$ are replaced by the values of $\psi(w) - \Psi_2(w)$ at the points $P$, then $\psi(w)$ is replaced by $\psi(w) - \Psi_2(w)$ and $M$ is unchanged.

There exists by the hypothesis on $\Psi_2(w)$ and by the use of an auxiliary theorem,* a polynomial $\omega(w)$ which takes on the same values as $\Psi_2(w)$ in the points $P$ and is such that

$$|\Psi_2(w) - \omega(w)| \leqq \frac{\epsilon}{2}, \text{ for } w \text{ on } B + C.$$

There results by means of (9.5) the inequality

$$|\Psi(w) - [\omega(w) + \omega''(w)]| \leqq M + \epsilon, \text{ for } w \text{ on } C,$$

* Walsh, these Transactions, vol. 30 (1928), pp. 307–332; Theorem X, p. 319. See also the proof of Theorem 7 of the present paper.

and the polynomial $\omega(w)+\omega''(w)$ satisfies the given auxiliary conditions for the $\omega_n(w)$. That is, the greatest lower bound of

$$|\Psi(w) - \pi(w)|, \text{ for } w \text{ on } C,$$

for all polynomials $\pi(w)$ satisfying the auxiliary conditions, is not greater than $M$, so we have

$$\lim_{n\to\infty} [\text{bound } |\Psi(w) - \omega_n(w)|, w \text{ on } C] = M,$$

and the desired result is a consequence of Theorem 14.

We mention explicitly that if an arbitrary function $\Psi_2(w)$ has no singularities other than in the infinite region of which $C$ is the boundary, then on $B+C$ the function $\Psi_2(w)$ can be uniformly approximated as closely as desired by a polynomial in $w$* and hence satisfies the conditions of Theorem 15.

It would be interesting to study the approximation on $C$ by polynomials of a function with singularities other than in $B$ and in the infinite region of which $C$ is the boundary; this problem arises only when $C$ separates the plane into more than two regions, and for the present the problem must remain unsolved.

There are evidently other questions that are left unanswered by the present treatment. Is it true† that if the region $B$ in Theorem 15 is a Jordan region the sequence $\{\omega_n(w)\}$ converges uniformly to the function $\psi(w)$ in the closed region? What can be said if in Theorem 15 the function $\Psi(w)$ is simply an arbitrary function continuous on $C$? What are the facts if in Theorem 14 we do not require the number of points $P$ to be finite? Under suitable restrictions much of the reasoning we have given can be extended with ease—for instance the proof of the existence of the minimizing function in Theorem 1 holds for approximation on $C$ to an arbitrary limited function (continuous or not) defined on $C$—but under other conditions it may happen (compare Nevanlinna, loc. cit., Abschnitt II) that the discussion given requires substantial modification.

**10. Existence and uniqueness of the Tchebycheff polynomial with auxiliary conditions.** For the situation of Theorem 15, the existence of the Tchebycheff polynomial with auxiliary conditions follows (in case the degree of the polynomial is so large that it is possible to make the polynomial satisfy the auxiliary conditions) from Montel's theory of normal families of functions. For the class of all polynomials $\omega_n(w)$ of a given degree $n$ satisfying the auxiliary conditions, the number

* Walsh, Mathematische Annalen, vol. 96 (1926), pp. 437–450; footnote, p. 441.

† The corresponding question is raised by Julia (loc. cit.) for the special problem treated by him, but is not answered.

$$\omega_n = \max | \Psi(w) - \omega_n(w) | , \text{ for } w \text{ on } C,$$

has a greatest lower bound $b$. There exists a sequence of polynomials $\omega_n^{(k)}(w)$ all of degree $n$ such that

$$\lim_{k\to\infty} [\max | \Psi(w) - \omega_n^{(k)}(w) | , \text{ for } w \text{ on } C] = b.$$

From this sequence can be extracted a subsequence which converges in $B$, uniformly on an arbitrary point set interior to $B$. The limit of this subsequence is then a polynomial $\omega_n(w)$ of degree $n$ which satisfies the auxiliary conditions, the convergence is uniform in $B+C$, and we have for this particular polynomial

$$b = \max | \Psi(w) - \omega_n(w) | , \text{ for } w \text{ on } C.$$

The existence of the Tchebycheff polynomial with auxiliary conditions can, however, be established under much more general conditions.* We do not dwell on this point, but turn to the proof of the uniqueness of the Tchebycheff polynomial.

It is worth remarking that the proofs we have given for the convergence properties of the sequence of Tchebycheff polynomials do not depend in any way on the uniqueness of those polynomials, so that the material about to be given is not logically necessary by way of justification.

*Let $f(z)$ be an arbitrary function of $z$ continuous on a closed limited point set $C$ containing at least $n+2$ points.† Then the Tchebycheff polynomial $\pi_n(z)$ of degree $n$ for approximation to $f(z)$ on $C$ with the auxiliary conditions*

$$\pi_n(\alpha_i) = \gamma_i \qquad (i = 1, 2, \cdots, k) \tag{10.1}$$

*is unique provided $n \geqq k-1$ and provided that $\gamma_i = f(\alpha_i)$ if $\alpha_i$ is a point of $C$.*

Special cases occur here if some of the points $\alpha_i$ coincide with each other, so that (10.1) is considered to restrict not merely $\pi_n(\alpha_i)$ but also one or more derivatives of the polynomial for the value $z=\alpha_i$; we use $k$ to denote the total number of *equations* (10.1), not the number of distinct points $\alpha_i$. It is not necessary for the truth of this theorem so to restrict the values of these derivatives of the Tchebycheff polynomials, if any are prescribed, that they coincide with the derivatives (if any exist) of $f(z)$, even if $\alpha_i$ is a point of $C$.

The proof which follows is, except for minor changes due to the introduction of the auxiliary conditions (10.1), precisely the proof given by Tonelli for the case $k=0$.‡ For that reason, we give the proof merely in outline and

* By the methods used by de la Vallée Poussin, *Approximation des Fonctions*, Paris, 1919, p. 75.

† The proof holds, with only minor modifications, if we have but $n+1$ points in $C$.

‡ Annali di Matematica, (3), vol. 15 (1908), pp. 108–113.

refer to Tonelli for the details.*

Let $\pi_n(z)$ denote a Tchebycheff polynomial of degree $n$, and consider the difference

$$f(z) - \pi_n(z) = Y(z),$$

which is likewise continuous on $C$ and has on $C$ a maximum modulus $\mu$. We shall prove first that *the number of distinct points of $C$ in which $Y(z)$ reaches its maximum absolute value is greater than $n-k+1$.*

We suppose the contrary, that $|Y(z)|$ reaches its maximum absolute value in the points $z_1, z_2, \cdots, z_\nu$ of $C$, where $\nu \leqq n-k+1$. It will be noticed that no point $z_i$ coincides with an $\alpha_j$. Let $P(z)$ be a polynomial of degree $n$ (i.e. in the sense of §5) such that

$$P(z_i) = Y(z_i) \qquad (i = 1,2, \cdots, \nu),$$

$$P(\alpha_i) = 0 \qquad (i = 1,2, \cdots, k),$$

where the second equation is to be interpreted as requiring the suitable derivatives of $P(z)$ to vanish if the $\alpha_i$ are not all distinct. Such a polynomial $P(z)$ of degree $n$ always exists.

Let an arbitrary positive $\epsilon < \mu/2$ be given. By the continuity of the functions $Y(z)$ and $P(z)$ on $C$, there exists a $\delta$ so that $|z'-z''| < \delta$ implies

$$|Y(z') - Y(z'')| < \epsilon, \quad |P(z') - P(z'')| < \epsilon.$$

Denote by $C'$ the point set common to $C$ and the circles

$$|z - z_i| < \delta, \tag{10.2}$$

and by $C''$ the complementary set with respect to $C$. On $C''$ the function $Y(z)$ has a maximum absolute value $\mu'' < \mu$:

$$|Y(z)| \leqq \mu''.$$

On $C$, the polynomial $P(z)$ has some maximum modulus $M$.

If the positive number $\omega$ is chosen less than $(\mu-\mu'')/(2M)$, we have

$$|\omega P(z)| < (\mu - \mu'')/2$$

for $z$ on $C$. If $\omega$ is chosen sufficiently small, still less than $(\mu-\mu'')/(2M)$, we have (Tonelli, loc. cit., pp. 110–111)

$$|Y(z) - \omega P(z)| < \mu' < \mu \tag{10.3}$$

for $z$ on $C'$; this inequality is proved first for each individual circle (10.2), where $\mu'$ depends on the circle, and hence follows for the entire point set $C'$.

* The present writer takes exception to the corresponding discussion given by Julia, in particular to the statement (loc. cit., pp. 291–292) "en tout point où $|\pi_n|$ atteint son maximum, $|Q_n|$ atteint aussi son maximum." The fact is of course true, but one cannot conclude *a priori* without further demonstration that $|Q_n|$ attains its maximum *in more than a single point* in which $|\pi_n|$ attains its maximum.

But we have for $z$ on $C''$

$$(10.4)\qquad |f(z) - [\pi_n(z) + \omega P(z)]| = |Y(z) - \omega P(z)| \leqq |Y(z)| + |\omega P(z)| \leqq \mu'' + \frac{\mu - \mu''}{2} = \frac{\mu + \mu''}{2} < \mu.$$

We have thus exhibited a new polynomial $\pi_n(z)+\omega P(z)$ of degree $n$ which satisfies the auxiliary conditions and which by (10.3) and (10.4) deviates from $f(z)$ on $C$ by a quantity in absolute value less than $\mu$, which contradicts the assumption that $\pi_n(z)$ is a Tchebycheff polynomial.

The uniqueness of the Tchebycheff polynomial will now be established from the fact just proved, that the difference $Y(z)$ reaches its maximum absolute value in more than $n - k+1$ points. Assume two Tchebycheff polynomials of degree $n$ to exist, $\pi_n(z)$ and $\pi_n'(z)$, and let $\mu$ have the same significance as before. Then we have for $z$ on $C$

$$(10.5)\qquad \left| f(z) - \frac{\pi_n(z) + \pi_n'(z)}{2} \right| \leqq \frac{|f(z) - \pi_n(z)|}{2} + \frac{|f(z) - \pi_n'(z)|}{2} \leqq \mu,$$

so that $[\pi_n(z)+\pi_n'(z)]/2$ is likewise a Tchebycheff polynomial of degree $n$ which satisfies the given auxiliary conditions. The left-hand member of (10.5) can equal $\mu$ for a particular value of $z$ only if *both* of the two terms in the next member have the value $\mu/2$, and indeed only if the two expressions

$$f(z) - \pi_n(z), \quad f(z) - \pi_n'(z)$$

are equal for this particular value of $z$. There are, as we have just shown, at least $n-k+2$ distinct values of $z$ (necessarily distinct from the $\alpha_i$) at which the left-hand member of (10.5) has the value $\mu$, so there are at least $n-k+2$ distinct values of $z$ different from the $\alpha_i$ at which $\pi_n(z)$ and $\pi_n'(z)$ are equal. The difference of these two polynomials has at least $n+2$ common roots, if possible multiple roots at the $\alpha_i$ are counted according to their multiplicities, so this difference vanishes identically, the polynomials $\pi_n(z)$ and $\pi_n'(z)$ are identical, and the theorem is completely established.

The simplest possible examples show that the theorem fails if there is omitted the restriction $\gamma_i=f(\alpha_i)$ when $\alpha_i$ is a point of $C$.

**11. Convergence and divergence of approximating sequence on the boundary.** We are not in a position to give a complete discussion of the convergence *on the circumference* $C$ of the sequence $\{F_n(z)\}$ of Theorem 2, but we shall give here certain positive and negative results concerning this convergence. We have already given (§4) an example to show that $\lim_{n\to\infty} F_n(z)=F(z)$ may fail at a single point of $C$, even when the functions $F_n(z)$

are analytic on as well as within $C$. At exceptional points it may occur that we have $\lim_{n\to\infty} F_n(z)=0$. Indeed, if we choose

$$F_n(z) = e^{n/[n^2(z-1)-1]+n/(n^2+1)},$$

we have fulfilled the requirements (in particular $F_n(0)=1$) of the functions considered in §4 with reference to Theorem 2, and in addition we have

$$\lim_{n\to\infty} F_n(z) = 0 \text{ for } z = 1.$$

We shall now give an example to show that *in Theorem 2 the equation* $\lim_{n\to\infty} F_n(z)=F(z)$ *may fail at every point of* $C$. We make use of the function considered in §4:

$$\Phi_n(z) = e^{1/[n(z-1)-1]+1/(n+1)},$$

so that we have $\Phi_n(1)=e^{-n/(n+1)}<e^{-1/2}$, $n>1$; $\Phi_n(-1)=e^{n/[(2n+1)(n+1)]}>1$. There exists a positive quantity $\theta_n$ such that for $|\theta|<\theta_n$ we have

$$|\Phi_n(e^{i\theta})| < e^{-1/2},$$

and for $|\theta-\pi|<\theta_n$ we have

$$|\Phi_n(e^{i\theta})| > 1.$$

Choose $N_n$ such that $1/N_n<\theta_n/(2\pi)$. We shall use an $N_n$th root of unity:

$$\omega_n = e^{2\pi i/N_n}.$$

Consider the sequence $\{F_n(z)\}$:

$$\Phi_2(\omega_2 z),\ \Phi_2(\omega_2^2 z),\ \cdots,\ \Phi_2(\omega_2^{N_2} z),$$
$$\Phi_3(\omega_3 z),\ \Phi_3(\omega_3^2 z),\ \cdots,\ \Phi_3(\omega_3^{N_3} z),$$
$$\cdots\cdots\cdots\cdots\cdots,$$

where the rows of functions $\Phi_n(\omega_n^i z)$ are taken in order. The conditions of Theorem 2 (in particular $F_n(0)=1$) are satisfied, the functions $F_n(z)$ are analytic even for $|z|=1$, yet $\lim_{n\to\infty} F_n(z)$ exists for no value of $z$ on $C$. Indeed if an arbitrary $z=z_0$ on $C$ is chosen and an arbitrary $N$, then there exist $N'$ and $N''$ such that

$$|F_{N'}(z_0)| < e^{-1/2}, \quad N' > N,$$
$$|F_{N''}(z_0)| > 1, \quad N'' > N.$$

We shall find it convenient to introduce some new terminology. Given the function $f(z)$ and the sequence of functions $\{f_n(z)\}$ defined on a certain point set $E$. Suppose that when an arbitrary $\epsilon>0$ is given there exists $N$ such that

$$|f_n(z)-f(z)|<\epsilon, \quad n\geqq N,$$

except perhaps on a point set $E_n$ of measure less than $\epsilon$. Then we shall say that the sequence $\{f_n(z)\}$ is *quasiconvergent* or *quasiconverges* to the function $f(z)$ on the given point set, and we write

$$\operatorname*{qlim}_{n\to\infty} f_n(z) = f(z).$$

Actual convergence need not take place—in fact it will appear later that the sequence of functions $F_n(z)$ defined above in terms of the functions $\Phi_n(\omega_n^i z)$ quasiconverges on the circumference $C$ to the function unity.

The notion just introduced is related to Weyl's *essentially uniform convergence.* The sequence $\phi_n(z)$ converges essentially uniformly on $E$ to the function $\phi(z)$ if, when an arbitrary $\epsilon>0$ is given, the sequence $\phi_n(z)$ converges uniformly on $E$ to $\phi(z)$ except perhaps on a point set of measure less than $\epsilon$.

*If the sequence $\{f_n(z)\}$ quasiconverges on $E$ to the function $f(z)$, then from every subsequence of the $\{f_n(z)\}$ can be extracted a new subsequence which converges on $E$ essentially uniformly to $f(z)$, and conversely, if from every subsequence of the functions $\{f_n(z)\}$ can be extracted a new subsequence which converges on $E$ essentially uniformly to $f(z)$, then the original sequence quasiconverges on $E$ to the function $f(z)$.*

We prove first the direct theorem. Let $\epsilon>0$ be given and let $N_1, N_2, \cdots$ be the numbers in the definition of quasiconvergence corresponding to the values $\epsilon/2, \epsilon/4, \epsilon/8, \cdots$. Let $f_{M_k}(z)$, $M_k \geqq N_k$, be a function of the given subsequence. Then the new subsequence is to be $\{f_{M_k}(z)\}$ and we have

$$|f_{M_1}(z) - f(z)| < \frac{\epsilon}{2}, \text{ except perhaps on a set } E_1 \text{ of measure } < \frac{\epsilon}{2},$$

$$|f_{M_2}(z) - f(z)| < \frac{\epsilon}{4}, \text{ except perhaps on a set } E_2 \text{ of measure } < \frac{\epsilon}{4},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots.$$

Thus the sequence $\{f_{M_k}(z)\}$ converges on $E$ uniformly to the function $f(z)$ except perhaps on the point set $E_{n+1}+E_{n+2}+\cdots$, whose measure is less than $\epsilon/2^n$.

Let us prove now the converse theorem. If the theorem is not true, there exists an $\epsilon>0$ and an infinity of indices $n_1, n_2, \cdots$, such that

$$|f_{n_k}(z) - f(z)| \geqq \epsilon$$

on a point set of measure not less than $\epsilon$. Then from the sequence $\{f_{n_k}(z)\}$ can be extracted no subsequence which converges on $E$ essentially uniformly to $f(z)$, which contradicts our hypothesis.

A theorem due to Egoroff* is of interest here.

* Paris Comptes Rendus, vol. 152 (1911), p. 244.

*If $E$ be a measurable set of finite measure, a necessary and sufficient condition that a sequence $\{f_n(z)\}$ of measurable functions should converge to a function $f(z)$ almost everywhere on $E$ is that the sequence $\{f_n(z)\}$ converge essentially uniformly to $f(z)$ on $E$.*

In the present paper we are concerned only with measurable functions. We combine Egoroff's theorem with the preceding one:

*If $E$ be a measurable set of finite measure, a necessary and sufficient condition that a sequence $\{f_n(z)\}$ of measurable functions should quasiconverge to $f(z)$ on $E$ is that from every subsequence of the $\{f_n(z)\}$ can be extracted a new subsequence which converges almost everywhere to $f(z)$ on $E$.*

We use the concept of quasiconvergence in a preliminary theorem:

*Let the single auxiliary condition of Theorem 2 be $F_n(0)=1$. Then we have for the sequence $\{F_n(z)\}$ of Theorem 2 in the sense of quasiconvergence on the circumference $C$,*

$$\operatorname*{qlim}_{n\to\infty} F_n(z) = 1.$$

Each function $F_n(z)$ is, at least for $n$ sufficiently large, uniformly bounded interior to $C$. It follows from a well known theorem due to Fatou that

$$\lim_{\rho\to 1} F_n(\rho e^{i\phi}), \quad z = \rho e^{i\phi},$$

exists for almost all values of $\phi$. These values are the values of $F_n(z)$ to which we refer in the theorem to be proved.

If $C'$ is the circle $|z|=\rho<1$, then we have Cauchy's integral

$$F_n(0) = 1 = \frac{1}{2\pi i}\int_{C'} \frac{F_n(z)}{z}dz = \frac{1}{2\pi}\int_{C'} F_n(z)d\phi,$$

for on $C'$ it is true that $z=\rho e^{i\phi}$, $dz = i\rho e^{i\phi}d\phi = iz\,d\phi$. If in this formula we take the limit as a sequence of values $\rho$ approaches unity, the function $F_n(z)$ on $C'$ approaches the limit $F_n(z)$ on $C$ almost everywhere (i.e. for all values of $\phi$ except at most a set of measure zero), and the functions $F_n(\rho e^{i\phi})$ are uniformly bounded with respect to $\rho$ and $\phi$ for each value of $n$. It follows from a theorem due to Lebesgue that we may integrate term by term:

$$1 = \frac{1}{2\pi}\int_C F_n(z)d\phi.$$

Our theorem is now intuitively obvious. The values $F_n(z)$ can be plotted in the complex plane. If the values are suitably weighted, their average is unity, for the integral taken over $C$ can be interpreted as an average. This

average lies on or within the smallest convex polygon which contains all the points $F_n(z)$, and lies on the polygon itself only in case the polygon degenerates. If $\lim_{n\to\infty}[\max|F_n(z)|]=1$, and if the average of the points $F_n(z)$ is always unity, then the values of $F_n(z)$ itself can differ but little from unity; if $n$ is sufficiently large this difference must be small except on a point set of small measure. In considering the average, values $F_n(z)$ which are taken on only on a set of points of zero measure may be neglected, and the theorem does not regard such values in the conclusion.

The intuitive proof just given can be expressed in more rigorous terms. We show that if an arbitrary $\eta>0$ be given, then $N$ exists such that

$$|F_n(z)-1|\leqq\eta, \text{ for } n\geqq N,$$

except perhaps for $z$ on a point set of $C$ of measure less than or equal to $\eta$. If this statement is not exact, then there exists a positive $\eta$ such that

$$|F_n(z)-1|>\eta$$

for $z$ on a point set $E_n$ of $C$ of measure greater than $\eta$ and for an infinity of subscripts $n$. We use our original notation $F_n(z)$ for the functions of this subsequence and shall reach a contradiction. Choose $\epsilon$ so small that $|F_n(z)|<1+\epsilon$ together with $|F_n(z)-1|>\eta$ implies

$$\Re[F_n(z)]<\eta'<1, \text{ or } \Re[1-F_n(z)]>\eta',$$

for some $\eta'>0$, where the symbol $\Re$ denotes the real part of the quantity which follows it. Such an $\epsilon$ exists, as is obvious geometrically; it is sufficient to choose $\epsilon<(1+\eta^2)^{1/2}-1$. Choose $\epsilon$ so small also that we have

$$\epsilon<\frac{\eta\eta'}{2\pi-\eta}.$$

On the one hand it follows that

$$\frac{1}{2\pi}\int_C[1-F_n(z)]d\phi=0,$$

and on the other hand that

$$\begin{aligned}\Re\left\{\frac{1}{2\pi}\int_C[1-F_n(z)]d\phi\right\}&\\ =\frac{1}{2\pi}\int_{E_n}\Re[1-F_n(z)]d\phi+\frac{1}{2\pi}\int_{C(E_n)}\Re[1-F_n(z)]d\phi&\\ \geqq\frac{\eta\eta'}{2\pi}-\frac{\epsilon(2\pi-\eta)}{2\pi}>0,&\end{aligned}$$

which is a contradiction. In the proof, $C(E_n)$ denotes the set complementary to $E_n$, whose measure is not greater than $2\pi-\eta$, and we have used the fact, obvious geometrically, that if $|F_n(z)|<1+\epsilon$, then $\Re[1-F_n(z)]>-\epsilon$.

Our preliminary theorem is now established, and leads to another preliminary theorem:

*Let the single auxiliary condition of Theorem 2 be $F_n(0)=1$. Then we have on the circumference $C$*

$$\lim_{n\to\infty}\int_0^{2\pi}|F_n(z)-1|\,d\phi=0.$$

Indeed we can easily establish the following:

*If the uniformly bounded sequence of measurable functions $\{F_n(z)\}$ quasi-converges to the function $F(z)$ on the point set $E$ of finite measure, then we have*

$$\lim_{n\to\infty}\int_E|F_n(z)-F(z)|^p|\,dz|=0,\quad p>0.$$

Assume $|F_n(z)|\leqq M$, and let $m$ be the measure of $E$. Then if an arbitrary positive $\epsilon$ be given, there exists $N$ so that we have, for $n\geqq N$,

$$|F_n(z)-F(z)|<\epsilon$$

except perhaps on a point set $E_n$ of measure less than $\epsilon$. Let $C(E_n)$ denote the set complementary (with respect to $E$) to $E_n$. There follows

$$\int_E|F_n(z)-F(z)|^p|\,dz|=\int_{C(E_n)}|F_n(z)-F(z)|^p|\,dz|+\int_{E_n}|F_n(z)-F(z)|^p|\,dz|<m\epsilon^p+2^pM^p\epsilon,$$

which approaches zero with $\epsilon$, and the theorem is proved.

It is interesting to notice that the single condition

$$\lim_{n\to\infty}\int_0^{2\pi}|F_n(z)-1|^p d\phi=0,\quad p>0,$$

for $z$ on $C$ implies the result, in the sense of quasiconvergence on $C$,

$$\operatorname*{qlim}_{n\to\infty}F_n(z)=1.$$

In fact, the corresponding condition in the general case

$$\lim_{n\to\infty}\int_C|F_n(z)-F(z)|^p d\phi=0,\quad p>0,$$

implies, in the sense of quasiconvergence on $C$,

$$\operatorname*{qlim}_{n\to\infty} F_n(z) = F(z).$$

Otherwise we should have, for some $\epsilon>0$ and for an infinity of indices $n$,

$$|F_n(z) - F(z)| \geqq \epsilon$$

on a point set of measure greater than or equal to $\epsilon$, and hence for those indices

$$\int_C |F_n(z) - F(z)|^p d\phi > \epsilon^{p+1},$$

which is impossible.

The results we have now established lead directly to the following:

*In Theorems 2 and 3 we have, in the sense of quasiconvergence on the circumference C,*

$$\operatorname*{qlim}_{n\to\infty} F_n(z) = F(z), \qquad \operatorname*{qlim}_{n\to\infty} f_n(z) = f(z).$$

*We have also*

$$\lim_{n\to\infty} \int_C |F_n(z) - F(z)|^p d\phi = 0, \quad \lim_{n\to\infty} \int_C |f_n(z) - f(z)|^p d\phi = 0, \quad p > 0.$$

Transform the functions $F(z)$ and $F_n(z)$ as in §5, using formulas (5.3), (5.4), (5.10), (5.11), repeated as often as necessary and in the proper order. In the last stage we have a situation of type

$$\lim_{n\to\infty} \phi_{ni}(z) = \phi_i(z), \quad |z| < 1,$$

where $\phi_i(z)$ is identically a constant $\phi_i$ of modulus $M$, and where the auxiliary conditions to which the $\phi_{ni}(z)$ are subjected are of the form $\phi_{ni}(z') = \phi_i$, where $z'$ is a particular point interior to $C$. At this stage the theorem is true, as is obvious by considering the functions $\phi_{ni}(z)/\phi_i$, and by transforming with the substitution $w = (z-z')/(1-\bar{z}'z)$. For these transformed functions, the theorem reduces to the special case already considered. Hence by transforming again by the inverses of the previous transformations, in the inverse order, we see that the result stated is valid for $F(z)$ and $F_n(z)$ and therefore for $f(z)$ and $f_n(z)$.

It is conceivable that there should be difficulty in this application, due to the fact that the normal to $C$ does not remain the normal to $C$ under transformations of the form $w = (z-z')/(1-\bar{z}'z)$. This difficulty is apparent rather than actual, however, for Fatou's theorem applies to approach to $C$ not

merely along the normal to $C$ but also in any way in a triangle interior (except for a single vertex) to $C$.*

The result on quasiconvergence just proved can, by conformal mapping, be extended to an arbitrary simply connected region $B$ in the situation of Theorem 14, provided that the boundary values of the functions $\psi_n(w)$ are suitably defined—let us say by normal approach to the boundary after $B$ is mapped onto a circle—and provided that the term *measure* on $C$ is also properly interpreted, perhaps by measure on the unit circle when $B$ is mapped onto its interior. If $B$ is the interior of a rectifiable Jordan curve, however, the boundary values can be taken by approach along the normal to $C$, and the definition of measure can be taken in terms of arc length on $C$. For a point set of zero measure on the curve $C$ corresponds to a point set of zero measure on the unit circle, and conversely.† Moreover, approach almost everywhere to the boundary $C$ of $B$ in a triangle in $B$ means approach almost everywhere to the boundary of the unit circle in a triangle in that unit circle, and reciprocally, for if $B$ is mapped onto the interior of the unit circle, the mapping is conformal almost everywhere on the boundary.‡ The result

$$\lim_{n\to\infty} \int_C |F_n(z) - F(z)|^p ds = 0, \quad p > 0,$$

also naturally persists under the present circumstances, that $B$ is bounded by a rectifiable Jordan curve $C$.

In the numerous cases we have considered of approximation by Tchebycheff polynomials on the boundary $C$ of a region $B$ in the $w$-plane, the question of boundary values of the approximating functions $F_n(z)$ is automatically answered. For the Tchebycheff polynomials are continuous on and within the boundary $C$, and under conformal mapping approach to the unit circle $\gamma$ in the $z$-plane corresponds to approach to the boundary $C$ in the $w$-plane. If $z$ approaches along the radius a point of $\gamma$ so that the corresponding point $w$ approaches an accessible point of $C$, then a unique limit for the corresponding functional values of the approximating functions obviously exists in the

* It is not intended to imply, nor is it essential to our argument, that whenever the limit for radial approach exists, then the limit in this new sense also exists. It is sufficient for our purpose that the limit in this new sense exists almost everywhere.

† This theorem is due to Lusin. See Lusin and Privaloff, Annales de l'École Normale Supérieure, (3), vol. 42 (1925), p. 156.

‡ Carathéodory, Schwarz Festschrift, Berlin, 1914, pp. 19–41; pp. 40–41. It is there shown that the existence of a tangent implies conformality on the boundary, but for a rectifiable curve a tangent is known to exist almost everywhere. The method of proof used here gives a proof of Denjoy's extension of Fatou's theorem, to regions $B$ bounded by arbitrary rectifiable Jordan curves. Compare Denjoy, Paris Comptes Rendus, vol. 168 (1919), p. 387.

$w$-plane and hence also in the $z$-plane. By Fatou's theorem, $w$ can fail to approach a limit when $z$ approaches $\gamma$ along a radius at most on a set of measure zero on $\gamma$, so almost everywhere on $C$ the actual boundary values of the Tchebycheff polynomials are the boundary values in the sense of Fatou in the $z$-plane, provided "almost everywhere" is interpreted in terms of measure on the unit circle $\gamma$ in the $z$-plane.

It is interesting to note that the result we have proved for quasiconvergence on the boundary $C$ implies convergence in the interior of $B$. We prove the general result in the $z$-plane, which implies the result also in the $w$-plane.

*If the sequence of functions $\{F_n(z)\}$ each analytic and bounded interior to $C$: $|z|=1$ is uniformly bounded almost everywhere on $C$ (that is, the sequence of boundary values taken on almost everywhere is uniformly bounded) and approaches on $C$ in the sense of quasiconvergence the function $F(z)$, the boundary values* (taken on almost everywhere on $C$) of a function $F(z)$ analytic and bounded interior to $C$, then we have*

$$\lim_{n\to\infty} F_n(z) = F(z)$$

*throughout the interior of $C$, uniformly on any closed point set interior to $C$.*

In this theorem, *boundary value* means naturally boundary value in the sense of Fatou.

We have by the boundedness of the individual functions $F_n(z)$ and $F(z)$ as above, by integrating first over the circle $C'$: $|z|=\rho<1$ and then allowing a sequence of values $\rho$ to approach unity,

$$F_n(z) - F(z) = \frac{1}{2\pi i}\int_C \frac{F_n(t) - F(t)}{t - z}dt.$$

The function $1/(t-z)$ is uniformly bounded in $t$ and $z$ if we have $|z| \leqq \rho<1$:

$$\left|\frac{1}{2\pi i}\int_C \frac{F_n(t) - F(t)}{t - z}dt\right| \leqq \frac{1}{2\pi i(1-\rho)}\int_C |F_n(t) - F(t)|\,d\phi;$$

the last integral approaches zero, so we have

$$\lim_{n\to\infty} F_n(z) = F(z)$$

interior to $C$, uniformly on any closed point set interior to $C$.

* The equation

$$\operatorname{qlim}_{n\to\infty} F_n(z) = F(z)$$

for $z$ on $C$ under the present hypothesis on the sequence $\{F_n(z)\}$ implies the fact that $F(z)$ is the boundary values of a function analytic and bounded interior to $C$.

In this same order of ideas we can prove the following:

*Let $f(z)$ be analytic and bounded for $|z|<1$:*

$$|f(z)| \leqq N.$$

*If we have almost everywhere (i.e. for almost all values of $\phi$)*

$$\overline{\operatorname{bound}}_{\rho\to 1} f(\rho e^{i\phi}) \leqq M, \quad \phi \text{ constant},$$

*then we have*

$$|f(z)| \leqq M \text{ for } |z| < 1.$$

We have Cauchy's integral

$$f(0) = \frac{1}{2\pi i}\int_{C'} \frac{f(z)dz}{z} = \frac{1}{2\pi}\int_{C'} f(z)d\phi,$$

where $C'$ is the circle $|z|=\rho<1$. If we take the limit as $\rho$ approaches unity, we have

$$f(0) = \frac{1}{2\pi}\int_{C} f(z)d\phi,$$

where this integral refers as usual to the boundary values of $f(z)$, namely $\lim_{\rho\to 1} f(\rho e^{i\phi})$; this limit is almost everywhere not greater than $M$ in absolute value, so we have

$$|f(0)| \leqq M.$$

Let $z_0$ be any other point interior to $C$. We transform by setting $w=(z-z_0)\cdot[1-\bar{z}_0 z]^{-1}$, so that $C$ is transformed into itself and $z_0$ into $w=0$. By Fatou's theorem, the boundary values of $f(z)$ in the $w$-plane for radial approach to $C$ are equal almost everywhere to the boundary values of $f(z)$ in the $z$-plane for radial approach to $C$. Then we have in the $w$-plane almost everywhere on $|w|=1$, for the boundary values obtained by radial approach, $|f(z)|\leqq M$, from which it follows that the functional value for $w=0$ is also not greater than $M$ in absolute value:

$$|f(z_0)| \leqq M,$$

and the proof is complete.

**12. Miscellaneous problems.** We shall consider in this section various problems, some solved and others unsolved, and make a few additional remarks relative to the problems already treated in detail.

12.1. As a first problem, we compare the two measures of approximation, least squares and that of Tchebycheff. For the sake of simplicity* we restrict

* The results on approximation by least squares hold with some changes also for much more general curves and functions; see Szegö, Mathematische Zeitschrift, vol. 9 (1921), pp. 218–270.

ourselves to the case of approximation on the unit circle $C$ to a rational function $r(z)$ with no singularities on $C$.

The least-squares method of approximation to $r(z)$ on $C$ by polynomials in $z$ is to study the sequence of polynomials $p_n(z)$ of respective degrees $n$ which make the integral

$$\int_C | r(z) - p_n(z) |^2 ds$$

a minimum. If we set $r(z) = r_1(z) + r_2(z)$, where $r_1(z)$ is analytic on and within $C$ and where $r_2(z)$ is analytic on and exterior to $C$ and vanishes at infinity, then the approximating polynomials $p_n(z)$ for approximation to $r(z)$ are the successive convergents of the Taylor expansion of $r_1(z)$:

$$r_1(z) = a_0 + a_1 z + a_2 z^2 + \cdots,$$
$$p_n(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n;$$

this determination of $p_n(z)$ follows from the formulas for the coefficients of $p_n(z)$:

$$a_n = \frac{1}{2\pi} \int_C r(z) \bar{z}^n ds = \frac{1}{2\pi i} \int_C r(z) \frac{dz}{z^{n+1}}.$$

If we approximate $r(z)$ on $C$ by means of polynomials $p_{-n}(z)$ in $1/z$ which vanish at infinity, then the method of least squares yields as the polynomials $p_{-n}(z)$ the successive convergents of the Taylor expansion at infinity of $r_2(z)$:

$$r_2(z) = a_{-1} z^{-1} + a_{-2} z^{-2} + \cdots,$$
$$p_{-n}(z) = a_{-1} z^{-1} + a_{-2} z^{-2} + \cdots + a_{-n} z^{-n}.$$

Thus we have the simple formulas $\lim_{n \to \infty} p_n(z) = r_1(z)$, for $|z|$ less than some $R > 1$, uniformly for $|z| \leqq R_1 < R$; $\lim_{n \to \infty} p_{-n}(z) = r_2(z)$, for $|z|$ greater than some $R' < 1$, uniformly for $|z| \geqq R_1' > R'$. There results from these two equations

$$r(z) = \lim_{n \to \infty} p_n(z) + \lim_{n \to \infty} p_{-n}(z),$$

uniformly for $z$ on $C$.

The circumstances are not nearly so simple for approximation by the method of Tchebycheff. Let $\{\tau_n(z)\}$ be the sequence of polynomials in $z$ for approximation on $C$ to $r(z)$, and $\{\tau_{-n}(z)\}$ the sequence of polynomials in $1/z$ for approximation on $C$ to $r(z)$, subject to the restriction* $\tau_{-n}(\infty) = 0$. We

* This restriction is not essential either here or above, but is a matter of convenience here, not merely to afford an analogy with the method of least squares, but to avoid complication in the present formulas. If the restriction is not made, the number $a_0$ enters in an extraneous manner.

now have, if $f(z)$ denotes the minimizing function of Theorem 3 for $r_2(z)$, $\lim_{n\to\infty} r_n(z) = f(z) + r_1(z)$, for $|z|$ less than some $R > 1$, uniformly for $|z| \leqq R_1 < R$; $\lim_{n\to\infty} r_{-n}(z) = f_1(z) + r_2(z)$, for $|z|$ greater than some $R' < 1$, uniformly for $|z| \geqq R_1' > R'$; here the function $f_1(z)$ corresponds to approximation to $r_1(z)$ on $C$ by functions analytic exterior to $C$, with the auxiliary condition $f_1(\infty) = 0$. Each of these two limits depends on both $r_1(z)$ and $r_2(z)$—of course $r_1(z)$ and $r_2(z)$ can be chosen independently of each other—and hence, in contrast to the limits with the method of least squares, there is no simple obvious relation connecting with $r(z)$ the sum of the two limits.

It would be interesting to study approximation in the sense of least squares of an arbitrary function (continuous or not) $f(z)$ defined on a rectifiable Jordan curve, to determine as we have done for simple functions in the corresponding case of Tchebycheff approximation, the properties of the sequences $\{p_n(z)\}$, $\{p_{-n}(z)\}$ defined as above.* It will be noticed that for the circle $C$ and for a function $f(z)$ analytic on $C$, the sequence $\{p_n(z)\}$ has the limit

$$\frac{1}{2\pi i}\int_C \frac{f(t)dt}{t - z}.$$

In the more general case, too, there is intimate connection with Cauchy's integral and with Fourier's series. The present writer hopes soon to publish some results on this general topic.

12.2. We consider now some of the cases of approximation already treated, modified so that the auxiliary conditions refer to points some of which may be outside of the point set on which the given function is approximated. To be sure, we have ordinarily considered approximation on the boundary $C$ of a region, and auxiliary conditions not on $C$ but interior to $C$, but if the functions involved have no singularities interior to $C$, this is not essentially different from approximation in the corresponding closed region with the same auxiliary conditions. In the present case the auxiliary conditions refer to points entirely without restriction as to location. The simplest situation is a generalization of Theorem 4:

*Let the function $f(z)$ be analytic on the closed limited point set $C$, not a single point, whose complementary set with respect to the entire plane is simply connected. Let $\{\pi_n(z)\}$ be the sequence of Tchebycheff polynomials for approximation to $f(z)$ on $C$ with the auxiliary conditions*

* Szegö (loc. cit., p. 239, footnote) also contrasts the two measures of approximation, least squares and that of Tchebycheff, but for less general cases: (1) where the given function is analytic in the closed region considered, (2) where the given function is $z^n$ and the approximating polynomial of degree $n-1$.

$$\pi_n(\alpha_i) = f(\alpha_i), \quad \alpha_i \text{ on } C \qquad (i = 1,2, \cdots, k),$$

$$\pi_n(\beta_i) = \gamma_i, \quad \beta_i \text{ not on } C \qquad (i = 1,2, \cdots, k').$$

*Let $C_R$ denote the largest curve of the family $C_\rho$ which contains within it no singular point of $f(z)$ and no point $\beta_i$ at which the assigned value $\gamma_i$ differs from the value $f(\beta_i)$ of the analytic extension of $f(z)$. Then the sequence $\{\pi_n(z)\}$ converges everywhere interior to $C_R$, uniformly on any closed point set interior to $C_R$; interior to $C_R$ the limit of the sequence is naturally the function $f(z)$ or an analytic extension of that function.*

In this theorem we require the suitable derivatives of $\pi_n(z)$ to coincide with those of $f(z)$, in case not all of the points $\alpha_i$ are distinct, and we also require the suitable derivatives $\pi_n^{(j)}(\beta_i)$ to coincide with the derivatives of $f^{(j)}(\beta_i)$ at multiple points $\beta_i$ interior to $C_R$.

Let us introduce the notation

$$\pi(z) = (z - \alpha_1) \cdots (z - \alpha_k)(z - \beta_1) \cdots (z - \beta_{k'}),$$

and let $p(z)$ be a polynomial of degree $k+k'-1$ such that

$$p(\alpha_i) = f(\alpha_i) \qquad (i = 1,2, \cdots, k),$$

$$p(\beta_i) = \gamma_i \qquad (i = 1,2, \cdots, k').$$

The function

$$F(z) = \frac{f(z) - p(z)}{\pi(z)}$$

is analytic everywhere within $C_R$, so by Theorem 4 there exist polynomials $P_n(z)$ of respective degrees $n$ such that

$$|F(z) - P_n(z)| \leqq \frac{M_1}{R_1^n}, \text{ for } z \text{ on } C, \quad R_1 < R.$$

It follows, since $\pi(z)$ is uniformly bounded on $C$, that we have

$$|[f(z) - p(z)] - \pi(z)P_n(z)| \leqq \frac{M_2}{R_1^n}, \text{ for } z \text{ on } C,$$

which we write in the form

$$|f(z) - [p(z) + \pi(z)P_n(z)]| \leqq \frac{M_2}{R_1^n}, \text{ for } z \text{ on } C.$$

The quantity in square brackets can be considered a polynomial of degree which we indicate by $m=n+k+k'$, so the inequality can be written

(12.1) $| f(z) - p_m(z) | \leqq \dfrac{M_3}{R_1^m}$, for $z$ on $C$, where $p_m(z) = p(z) + \pi(z)P_n(z)$.

To be sure, this inequality and the remark on the degree of $p_m(z)$ have been established not for all $m$ but merely for $m$ sufficiently large, but the introduction of suitable polynomials for the lacking values of $m$ presents no difficulty.

We have exhibited a set of polynomials which satisfy inequality (12.1) and for $m$ sufficiently large satisfy the prescribed auxiliary conditions; the Tchebycheff polynomials must therefore likewise satisfy this inequality for $m$ sufficiently large, so that the theorem follows from Theorem 4.

It will be noted that the sequence $\{\pi_n(z)\}$ can converge uniformly in no region $C_{R_1}$, with $R_1 > R$.

It is clear from the reasoning just given what modifications are to be made in Theorems 10, 13, 15, if auxiliary conditions for points exterior to $C$ are introduced; we leave the results to the reader. If auxiliary conditions for points exterior to $C$ are introduced in these theorems, there are of course no modifications in the conclusions of the theorems to be made so far as concerns the general regions $B$ not assumed bounded by analytic curves.*

12.3. Another problem related to those discussed in detail is that of approximation by functions other than rational polynomials. We take up the situation of Theorem 13 by way of illustration. Let $\{\chi_n(w)\}$ be a set of functions continuous on $C$; compare the latter part of §6. Let $F(w)$ be an arbitrary function continuous on $C$. The function $\Sigma_n(w)$ of the form

$$\Sigma_n'(w) = a_1\chi_1(w) + \cdots + a_n\chi_n(w) \tag{12.2}$$

such that

$$\max | \Sigma_n'(w) - F(w) | , \quad w \text{ on } C,$$

has the least possible value, $n$ being fixed, may be called the Tchebycheff polynomial $\Sigma_n(w)$ of order $n$ for approximation to $F(w)$ on $C$. When the functions $\Sigma_n(w)$ exist—this need not be the case—the questions naturally arise: What is

$$\lim_{n\to\infty} [\max | \Sigma_n(w) - F(w) | , \ w \text{ on } C] ? \tag{12.3}$$

What can be said of the convergence of the sequence $\{\Sigma_n(w)\}$ on $C$ and interior to $C$? What is the limit of this sequence?

Whenever the set $\{\chi_n(w)\}$ forms a *basis* for $B$, that is, if the functions $\chi_n(w)$ are analytic in $B$, continuous on $B+C$, and if an arbitrary function

* Compare Walsh, these Transactions, vol. 31 (1929), pp. 477–502, §7. This illustrates other restrictions on the sequence $\{\pi_n(z)\}$.

analytic on and within $C$ can be uniformly approximated on $B+C$ as closely as desired by a function of type (12.2), and if each function $\chi_n(w)$ can be so approximated on $B+C$ by a rational polynomial, then for any function $F(w)$ the limit (12.3) is the same for the functions $\Sigma_n(w)$ as for the rational Tchebycheff polynomials. In particular if $F(w)$ is the function $s(w)$ of Theorem 13, this limit is $M$ and we have

$$\lim_{n\to\infty} \Sigma_n(w) = g(w),$$

uniformly on any closed point set interior to $B$. This result obviously requires the existence of the Tchebycheff polynomial (12.2), but does not require its uniqueness. For special sets of functions $\{\chi_n(w)\}$, more explicit results are obtainable.

The study of a basis for a given region should be interesting also in the case that the given region is not simply connected. So far as the writer is aware, even the analogue of Theorem 1 has not been established in this case, although most of the reasoning for simply connected regions, with only minor modifications, would seem to carry over.

12.4. We mention a few other problems suggested by and related to the present discussion. Most obvious is perhaps the study of the rational Tchebycheff polynomial with auxiliary conditions, for approximation to a real function $f(x)$ on an interval or other linear point set $C$, where the auxiliary conditions refer to points of $C$ and also to points not belonging to $C$, where $f(x)$ is or is not continuous on $C$, and when the auxiliary conditions on $C$ do or do not coincide with the values of $f(x)$. Much of the discussion already given in the present paper applies to real polynomials, and many of the classical properties of the Tchebycheff polynomial can be readily extended to the new case.*

The present problems (e.g. of Theorem 10) should also be interesting if $C$ is a Jordan arc or more general point set, particularly where the given points $P$ lie on $C$. What can be said of the sequence of functions corresponding to the regions $C_\mu$? What occurs in Theorem 15 if some of the points $\alpha_i$ lie *on* $C$ *itself*?

Approximation is ordinarily considered, moreover, in the sense of addition or subtraction, as in the present paper, but related questions are easily framed. Let $f(z)$ be a given rational function of $z$. What function $\phi(z)$, perhaps satisfying certain auxiliary conditions, is such that

$$\overline{\text{bound}}\,|\,\phi(z)f(z)\,|$$

* Compare de la Vallée Poussin, loc. cit. In particular the proof of the existence of the Tchebycheff polynomial does not require the continuity of $f(x)$.

for $|z|\to 1$ is least? is such that this upper bound for $|z|\leqq 1$ is least? Does a polynomial $\phi_n(z)$ of given degree $n$ exist which satisfies the auxiliary conditions and for which this upper bound is unique? Is it true that $\lim_{n\to\infty} \phi_n(z)=\phi(z)$ uniformly for $|z|\leqq 1$? Some of the methods of the present paper serve to answer these questions and others, at least in part, and the writer hopes to consider these problems on another occasion.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# ON POWER CHARACTERS OF SINGULAR INTEGERS IN A PROPERLY IRREGULAR CYCLOTOMIC FIELD*

BY

H. S. VANDIVER

Let

$$\zeta = e^{2i\pi/l},$$

where $l$ is an odd prime and let $\mathfrak{p}$ be a prime ideal in the field $k(\zeta)$ and $\omega$ an integer in the field with $\mathfrak{p}$ and $\omega$ prime to $l$. Then there exists an integer $a$ such that, if $N(\mathfrak{p})$ is the norm of $\mathfrak{p}$,

$$\omega^{(N(\mathfrak{p})-1)/l} \equiv \zeta^a \pmod{\mathfrak{p}},$$

and we write

$$\{\omega/\mathfrak{p}\} = \left\{\frac{\omega}{\mathfrak{p}}\right\} = \zeta^a.$$

The symbol on the left is termed a power character. The integer $\omega$ is congruent to the $l$th power of an integer in the field $k(\zeta)$ modulo $\mathfrak{p}$ if, and only if,

$$\left\{\frac{\omega}{\mathfrak{p}}\right\} = 1.$$

If

$$\mathfrak{P} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_s$$

then

$$\left\{\frac{\omega}{\mathfrak{P}}\right\} = \left\{\frac{\omega}{\mathfrak{p}_1}\right\}\left\{\frac{\omega}{\mathfrak{p}_2}\right\} \cdots \left\{\frac{\omega}{\mathfrak{p}_s}\right\}.$$

For the case where $\omega$ is a unit in $k(\zeta)$ these power characters were considered by Kummer† who gave among other results the relation (1) below. Kummer also‡ gave the following theorem:

If $\mathfrak{p}$ and $\mathfrak{q}$ are different from each other and from $\lambda = (1-\zeta)$ and are both prime ideals in the regular cyclotomic field $k(\zeta)$, then

$$\left\{\frac{\mathfrak{p}}{\mathfrak{q}}\right\} = \left\{\frac{\mathfrak{q}}{\mathfrak{p}}\right\}.$$

The symbol $\{\mathfrak{p}/\mathfrak{q}\}$ is defined as follows:

$$\mathfrak{p}^{hh_1} = (\pi),$$

---

* Presented to the Society, August 29, 1929; received by the editors in January, 1930.

† Journal für Mathematik, vol. 44 (1852), pp. 121–130; Vandiver, Annals of Mathematics, vol. 30 (1929), pp. 487–491.

‡ Berichte der Akademie der Wissenschaften zu Berlin, 1858, 1859, 1861; Hilbert, *Bericht über die Theorie der Algebraischen Zahlkörper*, p. 471.

where $h\ h_1 \equiv 1 \pmod{l}$ and $\pi$ is an integer in $k(\zeta)$, $h$ being the class number of $k(\zeta)$; then

$$\{\mathfrak{p}/\mathfrak{q}\} = \left\{\frac{\mathfrak{p}}{\mathfrak{q}}\right\} = \left\{\frac{\pi}{\mathfrak{q}}\right\}.$$

Further $\pi$ is selected to be primary. A regular cyclotomic field is one in which $h$ is prime to $l$.

In 1909 Furtwängler proved that

$$\left\{\frac{\alpha}{\beta}\right\} = \left\{\frac{\beta}{\alpha}\right\}$$

where $\alpha$ and $\beta$ are integers in a field $K$ prime to each other and to $l$, at least one being primary, the said field $K$ including the field $k(\zeta)$.

Takagi* proved that the value of the symbol $\{\mu/\mathfrak{r}\}$ depends only upon the ideal class modulo $\mathfrak{f}$ to which the ideal $\mathfrak{r}$ belongs, where $\mathfrak{f}^{l-1}$ is the relative discriminant of the field $K$, with respect to $k$. In the above theorem $\mu$ is an integer in a field which contains $k(\zeta)$ and $\mathfrak{r}$ is an arbitrary ideal prime to $l$ in this field. Two ideals $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are said to belong to the same class modulo $\mathfrak{f}$ if, and only if, two integers $\gamma_1$ and $\gamma_2$ exist such that $\gamma_1\mathfrak{a}_1 = \gamma_2\mathfrak{a}_2$ with $\gamma_1 \equiv \gamma_2 \pmod{\mathfrak{f}}$. Other theorems concerning power characters have been given, but reference to the above will be sufficient to indicate the character of the results given in the present paper.

In the first place we may note that Furtwängler's result is not a direct generalization of Kummer's law of reciprocity, as the symbol $\{\mathfrak{p}/\mathfrak{q}\}$ is not employed in the general field; in fact it is not defined when $\mathfrak{p}$ belongs to an exponent which is a multiple of $l$. Furtwängler's result may be deduced from Takagi's, quoted above, the latter result obviously giving us information concerning the power character of $\{\mu/\mathfrak{r}\}$ when $\mathfrak{r}$ belongs to an exponent which is a multiple of $l$, as well as all other types of ideals.

The proofs of the results of Furtwängler and Takagi are extremely long and complicated. It seems to me possible that simpler proofs might be obtained for the cyclotomic field at least by introducing new points of view into the theory of the power characters in the cyclotomic field itself. Having this in view, in the present paper I shall show how the value of the symbol $\{\mu/\mathfrak{r}\}$ may be completely determined where $\mu$ is an integer and $\mathfrak{r}$ any ideal in the field $k(\zeta)$, the integer $\mu$ being singular and $k(\zeta)$ being properly irregular. A cyclotomic field $k(\zeta)$ is said to be properly irregular if it is irregular and none of the units $E_i$, $i = 1, 2, \cdots, (l-3)/2$, is the $l$th power of a unit in $k(\zeta)$, where $E_n = \epsilon^R$;

* Journal of the College of Science of Tokio University, 1922, §5, p. 39.

$$R=(1+sr^{-2n}+s^2r^{-4n}+\cdots+s^{(l-3)/2}r^{-n(l-3)})r^{l^2},$$

the symbol $s$ representing the substitution $(\zeta/\zeta^r)$ in the notation of the Kronecker-Hilbert symbolic powers; $r$ is a primitive root of $l$ and

$$\epsilon=\left(\frac{(1-\zeta^r)(1-\zeta^{-r})}{(1-\zeta)(1-\zeta^{-1})}\right)^{1/2}.$$

Also the integer $\omega$ in $k(\zeta)$ is said to be singular, if and only if $(\omega)$ is the $l$th power of an ideal $\mathfrak{j}$ in $k(\zeta)$ but $\mathfrak{j}$ is not a principal ideal.

In this type of field it will first be necessary to determine the value of the symbol $\{\omega/\mathfrak{p}\}$ where $\omega$ is a unit in $k(\zeta)$ and $\mathfrak{p}$ is an arbitrary ideal in $k(\zeta)$ prime to $l$. The germ of the method I shall employ to accomplish this is due to Kummer who, in his 1857 Memoir on Fermat's last theorem,* found the value of

$$\left\{\frac{E_n}{\mathfrak{p}}\right\},$$

where the class number of $k(\zeta)$ was divisible by $l$ and not by $l^2$ and $B_n$ was the single Bernoulli number in the set $B_1, B_2, \cdots, B_{(l-3)/2}$ which was divisible by $l$; $B_1=1/6$, $B_2=1/30$, etc.; $\mathfrak{p}$ was an arbitrary ideal in the field. However, during the course of this proof Kummer employed without proof or reference the relation

$$\frac{d_0^{ml^i}\log\phi(e^v)}{dv^{ml^i}}\equiv\frac{d_0^{ml^i}\log\phi_1(e^v)}{dv^{ml^i}}\pmod{l^{i+1}},$$

where $m$ is not a multiple of $l-1$ and the subscript zero of $d$ indicates that $v=0$ is substituted in the result after the operations are performed. Also $\phi$ and $\phi_1$ are polynomials in $e^v$ with rational integral coefficients and $\phi(\zeta)=\phi_1(\zeta)$.†

I encountered the principal difficulty in carrying through this investigation in connection with this unproved statement of Kummer; in fact, I have been able to establish it only under restrictions (cf. Lemma 2).

Concerning properly irregular cyclotomic fields, it should be noted that through some computations carried out recently by Miss E. T. Stafford and the writer it was found that the fields defined by $l=37, 59, 67, 101, 103,$

* Berlin Abhandlungen, Mathematisch-Physikalische Klasse, 1857, pp. 41–74.

† Several of the devices employed in my method for deriving this result, under restrictions, are not used by Kummer in any of his papers and I have no idea of the method he used if he really possessed a proof. It is possible that he employed the erroneous formula obtained by him and quoted on page 45, (7a), of the report on *Algebraic Numbers*, II, Bulletin, National Research Council, February, 1928, as differential coefficients, of the type used in the theorem under discussion, appear therein.

131, 149 and 157 are properly irregular. The work was carried through for all primes $l$ less than 200. No irregular fields other than properly irregular fields were discovered.

The theorem of Kummer's, given in the first article of his quoted in the present paper, is as follows:

$$\text{(1)} \qquad \text{ind } E_n \equiv \frac{r^{2n}-1}{2(1+a^{l-2n}-(a+1)^{l-2n})} \cdot \frac{d_0^{l-2n} \log \psi_a(e^v)}{dv^{l-2n}} \pmod{l};$$

where

$$\left\{\frac{E_n}{\mathfrak{q}}\right\} = \zeta^{\text{ind}E_n};$$

$a$ is a rational integer, $0<a<l-1$; $r$ is a primitive root of $l$,

$$\psi_a(x) = \sum_h x^{-(a+1)h+\text{ind}(g^h+1)},$$

$x$ is arbitrary, $\mathfrak{q}$ is an ideal prime in $k(\zeta)$ whose norm is $q^t$, $g$ is a primitive root of $\mathfrak{q}$, $h$ ranges over the integers $0, 1, 2, \cdots, q^t-2$, excepting $(q^t-1)/2$ if $q$ is odd and excepting zero if $q$ is even. The primitive root $g$ is selected so that

$$g^{(q^t-1)/l} \equiv \zeta \pmod{\mathfrak{q}}.$$

If

$$g^h + 1 \equiv g^k \pmod{\mathfrak{q}},$$

$0<k<q^t-1$, then we write

$$\text{ind } (g^h+1) = k.$$

The symbol

$$\frac{d_0^{l-2n} \log \psi_a(e^v)}{dv^{l-2n}}$$

means that the $(l-2n)$th derivative of $\log \psi_a(e^v)$ is taken with respect to $v$ and $v=0$ substituted in the result, $e$ being the Napierian base. Further $a$ is selected so that

$$1 + a^{l-2n} - (a+1)^{l-2n}$$

is prime to $l$. It is known that

$$\text{(2)} \qquad \psi_a(\zeta) = \prod \mathfrak{q}_c,$$

where $c$ ranges over the integers

$$cc_1 \equiv 1 \pmod{l},$$

and

$$|ac_1| + |c_1| > l,$$

$|x|$ is the least positive residue of $x$ modulo $l$, and $c_1$ is in the set $1, 2, \cdots, l-1$. Also, the $\mathfrak{q}_c$ represents the ideal obtained from $\mathfrak{q}$ by the substitution $(\zeta/\zeta^c)$. Assume that $\mathfrak{q}$ belongs to the exponent $jl^{i-1}$, where $j$ is prime to $l$ and $i>1$, so that $\mathfrak{g}^{jl^{i-1}(l-1)} = (\omega)$, where $\omega$ is an integer in $k(\zeta)$ such that

$\omega \equiv 1 \pmod{(1-\zeta)}$. Raising both sides of (2) to the power $jl^{i-1}(l-1)$ we obtain

$$(\psi_a(\zeta))^{jl^{i-1}(l-1)} = \eta \prod_c \omega_c,$$

where $\eta$ is a unit in $k(\zeta)$. Let $t$ be a positive integer such that $x^t\psi_a(x) = \psi_a'(x)$ is a polynomial in $x$. Then

$$\sum_h \zeta^{-(a+1)h+\operatorname{ind}(g^h+1)+t}$$

is a polynomial in $\zeta$, and if $\omega_c$ is obtained from $\omega$ by the substitution $(\zeta/\zeta^c)$ then

$$(\psi_a'(\zeta))^{jl^{i-1}(l-1)} = \eta \prod_c \omega_c. \tag{2a}$$

Now $\omega_c \equiv 1 \pmod{(1-\zeta)}$ and if

$$\omega_c = d_0 + d_1\zeta + \cdots + d_{l-2}\zeta^{l-2}$$

with the $d$'s rational integers, we write

$$\omega_c(x) = d_0 + d_1 x + \cdots + d_{l-2}x^{l-2} + \frac{1 - (d_0 + d_1 + \cdots + d_{l-2})}{l} \cdot \frac{x^l - 1}{x - 1}. \tag{2b}$$

Raise both sides of (2a) to the power $l$. We may then apply Lemma 2 (which is proved later) to the resulting relation, noting that $(\psi_a'(1))^{jl^i(l-1)} \equiv 1 \pmod{l^{i+1}}$; $\omega_c(\zeta) = \omega_c$; $\omega_c(1) = 1$ and if $m \not\equiv 0 \pmod{(l-1)}$ we obtain, after dividing through by $l$,

$$\begin{aligned}(l-1)jl^{i-1}\frac{d_0^{ml^i}\log\psi_a'(e^v)}{dv^{ml^i}} &\equiv (l-1)jl^{i-1}\frac{d_0^{ml^i}\log\psi(e^v)}{dv^{ml^i}} \\ &\equiv \sum_c \frac{d_0^{ml^i}\log\omega_c(e^v)}{dv^{ml^i}} \pmod{l^i},\end{aligned} \tag{3}$$

and using a relation which I have proved in another paper* we obtain

$$\sum_c \frac{d_0^{ml^i}\log\omega_c(e^v)}{dv^{ml^i}} \equiv \sum_c c^{ml^i}\frac{d_0^{ml^i}\log\omega(e^v)}{dv^{ml^i}} \pmod{l^i}. \tag{3a}$$

The expression $\sum_c c^{ml^i}$ may be reduced as follows. Consider

$$\sum_{h=1}^{l-1} \frac{|ah| + |h| - |(a+1)h|}{l} h^{ml^i} = L,$$

* These Transactions, vol. 31 (1929), p. 619, relation (3e).

which is readily seen to equal $\sum c_1^{ml^i}$ where $c_1$ ranges over those integers satisfying $|ac_1|+|c_1|>l$ in the set $1, 2, \cdots, l-1$. Now

$$|ah| = ah + lw,$$

where $w$ is a rational integer and

$$(ah)^{l^i m} \equiv |ah|^{l^i m} \pmod{l^{i+1}},$$

whence

$$h^{l^i m} \equiv (a)^{-l^i m + l^i(l-1)} |ah|^{l^i m} \pmod{l^{i+1}}.$$

Similarly we have

$$(h)^{l^i m} = |h|^{l^i m} \pmod{l^{i+1}}.$$

Hence

$$h^{l^i m} \equiv (a+1)^{l^i(l-1)-l^i m} |(a+1)h|^{l^i m} \pmod{l^{i+1}}.$$

Writing $b = l^i(l-1) - l^i m$, we obtain from the last two relations after dividing through by $l$ and noting that

$$\sum_h |ah| = 1 + 2 + \cdots + (l-1)$$

and similarly for $|h|$ and $|(a+1)h|$,

$$L \equiv \frac{a^b + 1 - (a+1)^b}{l} \sum_{n=1}^{l-1} n^{l^i m + 1} \pmod{l^i}.$$

Set

$$d = ml^i + 1.$$

We shall now prove the relation

$$S_d(l) \equiv (-1)^{d/2-1} l B_{d/2} \pmod{l^{i+2}},\ l > 3, \tag{3b}$$

where

$$S_d(l) = 1^d + 2^d + \cdots + (l-1)^d.$$

We have

$$S_d(l) = lb_d^* + l\binom{d}{2}(lb_{d-1}^*) + l^2 \frac{d(d-1)}{2 \cdot 3}(lb_{d-2}^*) + \cdots + l\frac{l^{r-1}}{r+1}\binom{d}{r}(lb_{d-r}^*) + \cdots,$$

where

$$b_{2a} = (-1)^{a-1} B_a,\ b_{2a+1}^* = 0,\ a > 0.$$

Consider the general term in this expansion. The numerator in the binomial coefficient is divisible by $l^i$. We shall now show that

$$\frac{l^{r-1}}{(r+1)!}$$

is a fraction whose numerator is divisible by $l$, for, by a known result, if $l^\mu$ is the highest power of the prime $l$ which divides factorial $r+1$ then

$$\mu = \frac{r+1-s}{l-1},$$

where $s=a_0+\cdots+a_n$ is the sum of the digits of $r+1$ to the base $l$,

$$r+1 = a_0 l^n + a_1 l^{n-1} + \cdots + a_n \qquad (0 \leqq a_1 < l).$$

Hence we have to show that

$$r-1 > \frac{r+1-s}{l-1}.$$

We have $(l-2)r>l-s$. Transposing $-r$ to the right hand member and also adding $1-l$ to each member, we obtain our results after dividing through by $l-1$. Hence (3b) follows if we note that the second term in the expansion is zero and $l>3$, and also by the von-Staudt-Clausen theorem $lb^*_{-r}$ is a fraction whose denominator is prime to $l$; and applying (3b) to the congruence involving $L$ we obtain

$$L \equiv ((a+1)^b - a^b - 1)(-1)^{d/2-1} B_{d/2} \pmod{p^i}.$$

We have

$$cc_1 \equiv 1 \pmod{l}.$$

Write

$$cc_1 = 1 + wl,$$

hence

$$\frac{1}{c^{ml^i}} \equiv (c_1)^{ml^i} \pmod{l^{i+1}},$$

and

$$c^{ml^i} \equiv (c_1)^b \pmod{l^{i+1}},$$

hence

$$\sum_m c^{ml^i} \equiv ((a+1)^{ml^i} - a^{ml^i} - 1)(-1)^{(b+1)/2} B_{(b+1)/2} \pmod{l^i}.$$

Consider (3a) and substitute the value which is obtained for $\sum_c c^{ml^i}$. We have

$$\sum_c \frac{d_0{}^{mi^i}\log\omega_c(e^v)}{dv^{mi^i}} \equiv ((a+1)^{i^i m} - a^{i^i m} - 1) \cdot(-1)^{(b+1)/2} B_{(b+1)/2} \frac{d_0{}^{mi^i}\log\omega(e^v)}{dv^{mi^i}} \pmod{l^i}. \tag{4}$$

The writer has proved† that

$$\frac{(n^i-1)\sum_{a=1}^{p-1} a^i}{p} = \sum_{a=1}^{p-1}\sum_{s=1}^{i}\binom{i}{s}\left(\frac{y_a}{a}\right)^s p^{s-1}a^i,$$

where $i$ is an arbitrary integer, $n$ is an integer prime to the prime $p$ and not unity, and

$$y_a \equiv -\frac{a}{p} \pmod{n} \quad (0 \leqq y_a < n).$$

Put $i=(b+1)$ in this relation; we obtain, as in the proof of (3b), if $p=l$

$$\frac{(n^{b+1}-1)\sum_{a=1}^{l-1} a^{b+1}}{l(b+1)} \equiv \sum_{a=1}^{l-1} y_a a^b \pmod{l^{i+1}}.$$

Using (3b) in this relation, we obtain

$$(n^{b+1}-1)(-1)^{(b-1)/2}B_{(b+1)/2} \equiv \sum_{a=1}^{l-1} y_a a^b \pmod{l^i}. \tag{5}$$

In a similar way we have, using (3b) and setting

$$s = l^{i-1}(l-1) - l^{i-1}m,$$

$$(n^{s+1}-1)(-1)^{(s-1)/2}B_{(s+1)/2} \equiv (s+1)\sum_{a=1}^{l-1} y_a a^s \pmod{l^i}.$$

Subtracting this relation from (5) we obtain

$$(-1)^{(b-1)/2}B_{(b+1)/2} \equiv \frac{(-1)^{(s-1)/2}B_{(s+1)/2}(n^{s+1}-1)}{(s+1)(n^{b+1}-1)} \pmod{l^i}.$$

Using (4) in connection with the above relation and noting that $n^{s+1}-1 \equiv n^{b+1}-1 \pmod{l^i}$ we obtain

$$\frac{d_0{}^{mi^i}\log\psi_a(e^v)}{dv^{mi^i}} \equiv \frac{a^{i^i m}+1-(a+1)^{i^i m}}{j} \cdot(-1)^{(s-1)/2}\frac{B_{(s+1)/2}d_0{}^{mi^i}\log\omega(e^v)}{l^{i-1}dv^{mi^i}} \pmod{l}. \tag{5a}$$

† Annals of Mathematics, vol. 18 (1917), p. 112, relation (7a).

In this relation set $i+1$ in place of $i$. Now if we put

$$D(\omega) = \frac{d\omega(e^v)}{dv}$$

then

$$\frac{d^{ml^{i+1}} \log \omega(e^v)}{dv^{ml^{i+1}}} = \frac{d^{ml^{i+1}-1}}{dv^{ml^{i+1}-1}} \left(\frac{D(\omega)}{\omega(e^v)}\right).$$

We shall consider

$$\frac{d^{ml^{i+1}-1}}{dv} \left(\omega(e^v)^{l^{i+1}} \cdot \frac{D(\omega)}{\omega(e^v)}\right).$$

Carrying out the differentiation of this fraction considered as a product, by Leibnitz's theorem, we have

$$\left[\frac{d^k(\omega(e^v)^{l^{i+1}})}{dv^k}\right]_{v=0} \equiv 0 \pmod{l^{i+1}},$$

and therefore, since $[\omega(e^v)]_{v=0} \equiv 1 \pmod{l}$,

$$\frac{d^{ml^{i+1}-1}}{dv^{ml^{i+1}-1}} \left[\frac{D(\omega)}{\omega(e^v)}\right]_{v=0} \equiv \left[\frac{d^{ml^{i+1}-1}}{dv} ((\omega(e^v))^{l^{i+1}-1} D(\omega))\right]_{v=0} \pmod{l^{i+1}}.$$

The expression in the large parenthesis on the right is a polynomial in $e^v$ with rational integral coefficients, say

$$f_0 + f_1 e^v + f_2 e^{2v} + \cdots + f_k e^{kv},$$

hence the right hand member may be written

$$f_1 + 2^{ml^{i+1}-1} f_2 + \cdots + k^{ml^{i+1}-1} f_k.$$

Hence

$$\left[\frac{d^{ml^i} \log \omega(e^v)}{dv^{ml^i}}\right]_{v=0}$$

is congruent (mod $l^{i+1}$) to

$$f_1 + 2^{ml^i-1} f_2 + \cdots + k^{ml^i-1} f_k,$$

and since if $s$ is a rational integer

$$s^{ml^i-1} \equiv s^{ml^{i+1}-1} \pmod{l^{i+1}},$$

we have

$$\frac{d_0^{ml^i} \log \omega(e^v)}{dv^{ml^i}} \equiv \frac{d_0^{ml^{i+1}} \log \omega(e^v)}{dv^{ml^{i+1}}} \pmod{l^{i+1}}.$$

In a similar way we find

$$\frac{d_0^{ml^i}\log\psi_a(e^v)}{dv^{ml^i}} \equiv \frac{d_0^{m}\log\psi_a(e^v)}{dv^{m}} \pmod{l}.$$

Using these relations together with (1) and (5a) we have the

THEOREM. *We have the relation*

$$\text{ind } E_n \equiv \frac{r^{2n}-1}{2j}(-1)^{(s-1)/2}\frac{B_{(s+1)/2}d_0^{(l-2n)l^i}\log\omega(e^v)}{l^i dv^{(l-2n)l^i}} \pmod{l} \tag{6}$$

*where*

$$\left\{\frac{E_n}{\mathfrak{q}}\right\} = \zeta^{\text{ind } E_n},$$

*$r$ is a primitive root of $l$, $\mathfrak{q}$ is a prime ideal in the field $k(\zeta)$ belonging to the exponent $jl^i$; $s=(2n-1)l^i$; $j$ prime to $l$;*

$$\mathfrak{q}^{jl^i(l-1)} = (\omega);$$

$$\omega \equiv 1 \pmod{\lambda};\ \lambda = (1-\zeta);\ i > 0;$$

*the $B$'s are the Bernoulli numbers, $B_1=1/6$, $B_2=1/30$ etc. The symbol*

$$\frac{d_0^a f(e^v)}{dv^a}$$

*indicates that $v=0$ is substituted in the $a$th derivative of $f(e^v)$ with respect to $v$, and $\omega(e^v)$ is defined as in* (2b).

The case where $i=0$ is disposed of in the first paper of Kummer's referred to in the present article. By means of these several results we may determine the value of the symbol $\{\eta/\mathfrak{q}\}$ where $\eta$ is any unit in $k(\zeta)$ and $\mathfrak{q}$ is any ideal prime to $l$ in $k(\zeta)$, since the units $e_i$, $i=1, 2, \cdots, (l-3)/2$, constitute an independent system of units, that is, there exists an integer $t$ such that $\eta$ can be expressed as the products of integral powers of the $E$'s. Since the field $k(\zeta)$ is properly irregular we may select $t$ to be prime to $l$. Hence (6) enables us to determine the value of the symbol.

Suppose now that $\theta$ is a singular integer in $k(\zeta)$ which is not necessarily a unit. Since the field is properly irregular, the second factor of the class number is prime to $l$ so that if $\mathfrak{q}$ is a prime ideal in $k(\zeta)$ prime to $\theta$ and $l$, then

$$(\mathfrak{q}\mathfrak{q}_{-1})^g = (\xi),$$

where $\xi$ is an integer in $k(\zeta)$ and $g$ is a rational integer prime to $l$. Hence by Furtwängler's law of reciprocity we have, if $\mathfrak{l}=(1-\zeta)$,

$$\left\{\frac{\theta}{qq_{-1}}\right\}^{\sigma} = \left\{\frac{\xi, \theta}{1}\right\}\left\{\frac{\xi}{\theta}\right\} = \left\{\frac{\xi, \theta}{\quad}\right\}. \tag{7}$$

To reduce the symbol

$$\left\{\frac{\theta}{qq_{-1}}\right\}$$

we note that this may be written as

$$\left\{\frac{\theta}{q}\right\}\left\{\frac{\theta_{-1}}{q}\right\}^{-1}.$$

But since

$$\theta\theta_{-1} = \eta\lambda^{l}$$

where $\eta$ is a unit in $k(\zeta)$ and $\lambda$ an integer in that field, we have

$$\theta_{-1} = \eta\lambda_1^{l}\theta^{l-1},$$

so that (7) gives

$$\left\{\frac{\theta}{q}\right\}^{2\sigma} = \left\{\frac{\xi, \theta}{1}\right\}\left\{\frac{\eta}{q}\right\}$$

and the last symbol on the right hand side is determined by means of (6); hence $\{\theta/q\}$ is completely determined.

We now consider

LEMMA I. *If $A_r$ represents the sum of terms of the type*

$$cu^{ap^r}v^{bp^r},$$

*where $a$, $b$ and $c$ are rational integers, $p$ is a prime integer, then*

$$(u + v)^{p^i s} = A_i + pA_{i-1} + p^2A_{i-2} + \cdots + p^iA_0,$$

*where $i$ and $s$ are rational integers.*

We may prove this result by means of the following theorem given by Kummer*:

The highest power $p^N$ of $p$ dividing the binomial coefficient

$$\frac{(A + B)!}{A!B!}$$

is given by

$$N = \epsilon_0 + \epsilon_1 + \cdots + \epsilon_l,$$

* Journal für Mathematik, vol. 9 (1832), p. 73; cf. also Stern, vol. 12 (1834), p. 288; Dickson, *History of the Theory of Numbers*, vol. I, p. 270–71; proof of Kummer corrected by Mitchell, these Transactions, vol. 17 (1916), pp. 170–72.

where the $\epsilon$'s are determined as follows. Set

$$A = a_0 + a_1 p + \cdots + a_l p^l,$$
$$B = b_0 + b_1 p + \cdots + b_l p^l,$$

where the $A$'s and $B$'s belong to the set $0, 1, \cdots, p-1$. We may find $c_1$ in this set and $\epsilon_i = 0$ or $1$ such that

$$a_0 + b_0 = \epsilon_0 p + c_0,$$
$$\epsilon_0 + a_1 + b_1 = \epsilon_1 p + c_1,$$
$$\epsilon_1 + a_2 + b_2 = \epsilon_2 p + c_2.$$

Whence

$$A + B = c_0 + c_1 p + \cdots + c_l p^l + \epsilon_l p^{l+1}.$$

Applying Kummer's theorem to Lemma I, let $d_0 \not\equiv 0$, and

$$p^i s = d_0 p^i + d_1 p^{i+1} + \cdots + d_j p^{i+j},$$

the $d$'s in the set $0, 1, \cdots, p-1$. Consider $p^r t$, with

$$t = e_0 + e_1 p + \cdots + e_k p^k, \quad e_0 \not\equiv 0,$$

the $e$'s in the set $0, 1, \cdots, p-1$. Then in

$$\frac{(A+B)!}{A!B!}$$

we have

$$A = p^r t,$$

and we also have

$$B = p^i s - p^r t$$

or

$$B = p^r(p - e_0) + p^{r+1}(p - (e_1 + 1)) + \cdots + p^{i-1}(p - (e_{i-r-1} + 1)) + E p^i$$

where $E$ is a rational integer $> 0$. Hence by Kummer's result

$$\frac{A+B}{A!B!}$$

is divisible by

$$p^{\epsilon_r + \cdots + \epsilon_i + \cdots} = p^{i-r+h},$$

where $h$ is a rational integer $\geqq 0$, since

$$\epsilon_r = \epsilon_{r+1} = \cdots = \epsilon_i = 1.$$

Kummer's argument for the proof of the theorem on which the above proof depends is very complicated, however, so I shall give another proof of Lemma I.

We have

$$(u+v)^p = A_1 + pA_0,$$
$$(u+v)^{p^2} = (A_1 + pA_0)^p = A_2' + pA_1' + p^2A_0',$$

where the $(A')$'s, $(A'')$'s have similar properties to the $A$'s. Now assume the theorem is true for the exponent $p^i$. We shall then prove it true for the exponent $p^{i+1}$. Raising both sides of the equation mentioned in the lemma for $s=1$, to the power $p$, obviously the result may be written

$$(A_i + pA_{i-1} + \cdots + p^{i-1}A_1)^p + p^{i+1}A_0'' .$$

Also, the term involving the parenthesis may be written

$$(A_i + pA_{i-1} + \cdots + p^{i-2}A_2)^p + p^iA_1'' .$$

Proceeding in this manner the lemma is proved for the exponent $p^{i+1}$, hence is true for any exponent which is a power of $p$. Using this expression for $(u+v)^{p^i}$ and raising to the power $s$ the result may be written

$$(A_i + pA_{i-1} + \cdots + p^{i-1}A_1)^s + p^iA_0.$$

Proceeding in this manner the lemma is established.

We now prove

LEMMA II. *Suppose*

$$\alpha = e^{2i\pi/p};$$

*p an odd prime*;

$$\phi = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_d\alpha^d$$

*where the a's are rational integers, and*

$$\phi_1 = a_0' + a_1'\alpha + a_2'\alpha^2 + \cdots + a_d'\alpha^d,$$

*d an arbitrary positive rational integer, the a"s rational integers*;

$$\phi = \phi_1,$$
$$\phi(x) = a_0 + a_1x + a_2x^2 + \cdots + a_dx^d,$$

*and similarly for* $\phi_1(x)$; *i is a rational integer* $\geqq 0$;

$$\phi(1) \equiv \phi_1(1) \pmod{p^{i+1}};$$
$$\phi_1(1) \not\equiv 0 \pmod{p};$$

*m a positive integer such that* $m \not\equiv 0 \pmod{p-1}$. *Then*

$$\frac{d_0^{mp^i} \log \phi(e^v)}{dv^{mp^i}} \equiv \frac{d_0^{mp^i} \log \phi_1(e^v)}{dv^{mp^i}} \pmod{p^{i+1}}.$$

As shown in a previous article (these Transactions, 1929, pp. 617–8) we obtain from $\phi=\phi_1$ the following identity in $e^v$:

$$\phi(e^v) = \phi_1(e^v) + (e^{vp} - 1)V + \frac{e^{vp} - 1}{e^v - 1}c_1,$$

where $c_1$ is a rational integer. Setting $v=0$ in this relation we obtain, since

$$\phi(1) \equiv \phi_1(1) \pmod{p^{i+1}},$$

the relation

$$pc_1 \equiv 0 \pmod{p^{i+1}},$$

whence

$$c_1 = cp^i,$$

where $c$ is a rational integer. The resulting relation may be written in the form

$$\frac{\phi(e^v)}{\phi_1(e^v)} = 1 + (e^{vp} - 1)\frac{V}{\phi_1(e^v)} + cp^i\frac{e^{vp} - 1}{\phi_1(e^v)(e^v - 1)}. \tag{8}$$

Set

$$y = (e^{vp} - 1)W + \frac{e^{vp} - 1}{e^v - 1}Z,$$

where

$$W = \frac{V}{\phi_1(e^v)}$$

and

$$Z = \frac{cp^i}{\phi_1(e^v)}.$$

We have

$$\frac{d\log(1+y)}{dv} = \frac{dy/dv}{1+y}$$

and

$$\begin{aligned}\frac{d^{mp^i}\log(1+y)}{dv^{mp^i}} &= \frac{d^{mp^i-1}}{dv^{mp^i-1}}\left(\frac{1}{1+y}\cdot\frac{dy}{dv}\right)\\ &= \frac{1}{1+y}\,\frac{d^{mp^i}y}{dv^{mp^i}} + (mp^i - 1)\frac{d}{dv}\left(\frac{1}{1+y}\right)\cdot\frac{d^{mp^i-1}y}{dv^{mp^i-1}}\\ &\quad + \cdots + \frac{d^{mp^i-1}}{dv^{mp^i-1}}\left(\frac{1}{1+y}\right)\cdot\frac{dy}{dv}.\end{aligned}$$

Now

$$\frac{d_0 y}{dv} \equiv 0 \pmod{p},$$

and if $k \not\equiv 0$,

$$\left[\frac{d^k}{dv^k}\left(\frac{1}{1+y}\right)\frac{d^{mp^i-k}y}{dv^{mp^i-k}}\right]_{v=0} \equiv \left[\frac{d^k}{dv^k}\left(\frac{1}{1+y}\right)\frac{d^{mp^i-k}x}{dv^{mp^i-k}}\right]_{v=0} \pmod{p^{i+1}},$$

if

$$x = (e^{vp} - 1)W.$$

Also

$$\left[\frac{d^k}{dv^k}\left(\frac{1}{1+y}\right)\right]_{v=0} \equiv \left[\frac{d^k}{dv^k}\left(\frac{1}{1+x}\right)\right]_{v=0} \pmod{p^i}.$$

Now if $k=0$, we note that

$$\frac{d_0^{mp^i} y}{dv^{mp^i}} \equiv \frac{d_0^{mp^i} x}{dv^{mp^i}} \pmod{p^{i+1}}$$

since $m \not\equiv 0 \pmod{(p-1)}$. Hence

$$\frac{d_0^{mp^i} \log(1+y)}{dv^{mp^i}} \equiv \frac{d_0^{mp^i} \log(1+x)}{dv^{mp^i}} \pmod{p^{i+1}}.$$

We have

$$\frac{d \log(1+x)}{dv} = \frac{dx/dv}{1+x},$$

$$\frac{d^{mp^i} \log(1+x)}{dv^{mp^i}} = \frac{d^{mp^i-1}}{dv}\left(\frac{dx/dv}{1+x}\right),$$

whence

$$\frac{dx/dv}{1+x}(1+x)^{p^i} = \frac{dx}{dv}(1+x)^{p^i-1}.$$

Differentiating the left hand member as a product we obtain

$$\frac{d^{mp^i} \log(1+x)}{dv^{mp^i}} \equiv \frac{d^{mp^i-1}}{dv}\left(\frac{dx/dv}{1+x}\right) \equiv \frac{d^{mp^i-1}}{dv}\left[\frac{dx}{dv}(1+x)^{p^i-1}\right]$$

(mod $p^{i+1}$) if $v=0$, since

$$\left[\frac{dx}{dv}\right]_{v=0} \equiv 0 \pmod{p}$$

and

$$[(1+x)^{p^i}]_{v=0} \equiv 1 \pmod{p^{i+1}}.$$

We also have

$$(1+x)^{p^i-1} = 1 + (p^i-1)x + \binom{p^i-1}{2}x^2 + \cdots + x^{p^i-1},$$

$$(p^i-1)(p^i-2)\cdots(p^i-r) \equiv (-1)^r r! \pmod{p^i}.$$

Hence

$$(1+x)^{p^i-1} \equiv 1 - x + x^2 - \cdots + x^{p^i-1} \pmod{p^i}$$

and since $dx/dv$ for $v=0$ is divisible by $p$ we have

$$\frac{d^{mp^{i-1}}}{dv}\left[\frac{dx}{dv}(1+x)^{p^{i-1}}\right]_{v=0} \equiv \frac{d^{mp^{i-1}}}{dv}\left[\frac{dx}{dv} - x\frac{dx}{dv} + x^2\frac{dx}{dv} - \cdots + \frac{dx}{dv}x^{p^i-1}\right]_{v=0} \pmod{p^{i+1}}.$$

Since

$$\frac{1}{n}\frac{dx^n}{dv} = x^{n-1}\frac{dx}{dv}$$

it follows that the right hand member of the above reduces to

$$\left[(-1)^{n-1}\sum_{n=1}^{p^i}\frac{1}{n}\frac{d^{mp^i}x^n}{dv}\right]_{v=0}.$$

Now take logarithms of (8) and differentiate $mp^i$ times. We then obtain, from what precedes,

$$\left[(-1)^{n-1}\sum_{n=1}^{p^i}\frac{1}{n}\frac{d^{mp^i}x^n}{dv}\right]_{v=0} \equiv \frac{d_0^{mp^i}\log\phi(e^v)}{dv^{mp^i}} - \frac{d_0^{mp^i}\log\phi_1(e^v)}{dv^{mp^i}} \pmod{p^{i+1}}.$$

It will now be shown that the left hand member of the above is congruent to zero (mod $p^{i+1}$). We have

$$x^n = ((e^{vp}-1)W)^n.$$

Assume that $n$ is prime to $p$.

Then

$$\frac{d^{p^im}(W_1(e^{vp}-1))}{dv^{p^im}} = (e^{vp}-1)\frac{d^{p^im}W_1}{dv^{mp^i}} + p^im\frac{d(e^{vp}-1)}{dv}\frac{d^{p^im-1}W_1}{dv^{p^im-1}} + \cdots + \frac{W_1 d^{p^im}(e^{vp}-1)}{dv^{p^im}}, \tag{9}$$

where

$$W_1(e^{vp}-1) = W^n(e^{vp}-1)^n.$$

We have obviously

$$\frac{d^k(e^{vp}-1)}{dv^k} = p^k e^{vp}. \tag{10}$$

Consider the general term in the above expression, namely

$$\binom{p^i m}{j} \frac{d^j(e^{vp}-1)}{dv^j} \frac{d^{p^i m-j} W_1}{dv^{p^i m-j}}.$$

According to Lemma I,

$$\binom{p^i m}{j}$$

is divisible by $p^{i-s}$, if $j=qp^s$; $s \geqq 0$. Using (10) for $v=0$ the general term becomes a fraction whose denominator is prime to $p$ and whose numerator is divisible by $p^{i-s+qp^s}$. Hence it is divisible by $p^{i+1}$ since $i-s+qp^s \geqq i+1$. Consequently our term $x^n/n$ in the original expansion becomes after the differentiation and substitution of $v=0$ a fraction whose denominator is prime to $p$ and whose numerator is divisible by $p^{i+1}$. Now consider the term in the original expansion $x^{p^a h}/p^a h$ where $p^a h \leqq p^i m$ and $h$ is prime to $p$. Put the expression $(W(e^{vp}-1))^{p^a h}$ in the form $W_2(e^{vp}-1)^{p^a}$. We have

$$(e^{vp}-1)^{p^a} = e^{vp^{a+1}} - p^a e^{vp(p^a-1)} + \binom{p^a}{2} e^{vp(p^a-2)} - \cdots - 1.$$

Consider the general term of this expansion, namely

$$\pm \binom{p^a}{j} e^{vp(p^a-j)},$$

where $j=qp^r$, $q$ being prime to $p$. Also consider from this the general term obtained in

$$\frac{d^s(e^{vp}-1)^{p^a}}{dv^s},$$

that is,

$$\pm p^s \binom{p^a}{j} (p^a-j)^s e^{vp(p^a-j)}.$$

As before the expression

$$\binom{p^a}{j}$$

is divisible by $p^{a-r}$. Hence this general term becomes after substituting $v=0$ an integer divisible by

$$p^{a-r} \cdot p^s \cdot (p^a - qp^r)^s,$$

or is divisible by

$$p^{a-r+rs+s}.$$

This exponent is obviously greater than or equal to $a+s$ for $s>0$ so that we have, for $s \geqq 0$,

$$\frac{d_0^s(e^{vp}-1)^{p^a}}{dv^s} \equiv 0 \pmod{p^{a+s}}. \tag{11}$$

Now consider

$$\frac{d^{p^i m}(W_2(e^{vp}-1)^{p^a})}{dv^{p^i m}}.$$

Expand by Leibnitz's theorem and consider the general term:

$$\binom{p^i m}{p^k n} \frac{d^{p^k n}(e^{vp}-1)^{p^a}}{dv^{p^k n}} \cdot \frac{d^{p^i m - p^k n} W_2}{dv^{p^i m - p^k n}}$$

when $k \geqq 0$. For $v=0$ it is divisible by

$$p^{i-k+a+p^k n}$$

which is divisible by $p^{i+1+a}$ since $h$ is prime to $p$ and $p^k n - k \geqq 1$. Hence all terms in the original expansion of the right hand member of (8), after differentiation $p^i m$ times and substituting $v=0$, become fractions whose denominators are prime to $p$ and whose numerators are all divisible by $p^{i+1}$. Hence the lemma is established.

We shall close this article by extending to higher powers of $l$ the criterion of Kummer and Takagi* that a principal ideal be the $l$th power of an ideal in $k(\zeta)$. Assume that

$$(\omega) = \mathfrak{a}^{l^i},$$

where $\mathfrak{a}$ is an ideal in $k(\zeta)$ and $\omega$ is an integer in that field. Hence from (2) and the formulas following it,

$$\prod_c \omega_c^{l-1} = \zeta^k \theta^{l^i},$$

where $\theta \equiv 1 \pmod{\lambda}$; $\lambda = (1-\zeta)$; also $cc_1 \equiv 1 \pmod{l}$ and $c_1$ ranges over the integers which satisfy $|ac_1| + |c_1| > 1$. Replace this relation by means of an identity in $e^v$ as usual and differentiate $ml^i$ times; apply Lemma II and substitute $v=0$. We obtain, as in the proof of our main theorem,

$$B_{(s+1)/2} \frac{d_0^{(l-2n)l^i} \log \omega(e^v)}{dv^{(l-2n)l^i}} \equiv 0 \pmod{l^{i+1}},$$

where $s=(2n-1)l^i$.

* Kummer, Berlin Abhandlungen, Mathematisch-Physikalische Klasse, 1857, p. 65; Takagi, Proceedings of the Physico-Mathematical Society of Japan, (3), vol. 4 (1922), pp. 170–182; Journal für die reine und angewandte Mathematik, vol. 157 (1927), pp. 230–8.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

# ON GIBBS'S PHENOMENON FOR THE DEVELOPMENTS IN BESSEL'S FUNCTIONS*

BY

CHARLES N. MOORE

**Introduction.** Several investigations have recently been published† concerning Gibbs's phenomenon in the case of the developments in Bessel's functions. As we should expect, the behavior of the developments in the neighborhood of the origin presented special difficulty, and the investigators limited themselves to very special types of functions‡ in dealing with this particular situation. It is the purpose of the present note to point out that the asymptotic formulas§ for the coefficients of the developments in Bessel's functions that have been obtained in two of my previous papers on this subject furnish a simple method of attacking the point in question. From these formulas the nature of Gibbs's phenomenon at the origin can be readily derived for general classes of functions. The same formulas also serve to establish the facts with regard to Gibbs's phenomenon at interior points of the interval (0, 1) and at the end point 1.

**1. The cause of Gibbs's phenomenon at the origin.** I will illustrate the application of these formulas to the discussion of Gibbs's phenomenon by dealing with the point $x=0$ and the class of functions considered in the earlier paper. This class includes functions which in the interval (0, 1) have a second derivative that is finite and integrable, except perhaps for a finite number of points at which the function itself or its first derivative has a finite jump. For such functions we have¶ for the coefficient of the general term of the development in Bessel's functions the following asymptotic

---

* Presented to the Society, April 7, 1928, and November 30, 1929; received by the editors in September, 1929, and February, 1930.

† Cf. R. G. Cooke, Proceedings of the London Mathematical Society, (2), vol. 27 (1928), pp. 171–192; J. R. Wilton, Journal für die reine und angewandte Mathematik, vol. 159 (1928), pp. 144–153.

‡ Cooke considers only the case where the function developed is a constant in the interval $(0 \leqq x \leqq 1)$; Wilton requires that in the neighborhood of the origin the function developed be of the form $l+Ax^{\nu+1/2}+o(x^{\nu+1/2})$, where $l$ and $A$ are constants and $\nu$ is the index of the Bessel's function, and he further restricts his discussion to the case where $\nu$ does not exceed 3/2.

§ Cf. these Transactions, vol. 12 (1911), p. 183, formula (8); vol. 21 (1920), p. 110, formula (15). The papers quoted here will be subsequently referred to as Transactions II and Transactions III respectively. An earlier paper (vol. 10 (1909), pp. 391–435) will be referred to as Transactions I.

¶ Loc. cit., Transactions II.

expression:

$$(1)\quad \sum_{m=1}^{m=k}\frac{M_m}{n^{1/2}}\cos\left[(n\pi+q)c_m-\frac{1}{4}(2\nu+3)\pi\right]$$
$$+\frac{Mf(1)}{n^{1/2}}\cos\left[n\pi+q-\frac{1}{4}(2\nu+3)\pi\right]+\frac{\nu Kf(0)}{\lambda_n}+r_n,$$

where the $c$'s are the points of discontinuity of the function developed, $r_n$ is the general term of an absolutely convergent series, and the $M$'s, $K$, and $q$ are constants. The precise value of $K$ is readily determined. From (34) and (29) of Transactions II we have

$$K'=\int_0^\infty J_{\nu+1}(y)dy=1.$$

From (38) of Transactions II we have $K=2K'/A^2=2/(2/\pi)=\pi$, since from (44) of Transactions I we have $A=(2/\pi)^{1/2}$.

It was shown in Lemma 2 of Transactions II that the first and second terms of (1), multiplied by $J_\nu(\lambda_n x)$, yield series uniformly convergent in the neighborhood of the origin, and the same is obviously true of the fourth term. Hence the Gibbs's phenomenon cannot occur if the third term is missing, and therefore there will be none in case $\nu=0$ or $f(0)=0$. If neither of these conditions holds, the presence of Gibbs's phenomenon can be readily inferred from the lemma of the following section.

2. We prove the following lemma:

LEMMA. *The series whose general term is*

$$(2)\qquad a_n=\int_{\lambda_n}^{\lambda_{n+1}}\frac{J_\nu(xu)}{u}du-\frac{J_\nu(\lambda_n x)}{\lambda_n}(\lambda_{n+1}-\lambda_n)$$

*is absolutely and uniformly convergent in the interval* $(0\leqq x\leqq 1)$.

We have

$$(3)\qquad a_n=\int_{\lambda_n}^{\lambda_{n+1}}\left[\frac{J_\nu(xu)}{u}-\frac{J_\nu(\lambda_n x)}{\lambda_n}\right]du.$$

By the law of the mean the integrand in (3) is equal to

$$(u-\lambda_n)\left[\frac{uxJ_\nu'(xu)-J_\nu(xu)}{u^2}\right]\qquad(u=\lambda_n+\theta,\ 0<\theta<\lambda_{n+1}-\lambda_n)$$

$$(4)\qquad=(u-\lambda_n)\left[\frac{(\lambda_n+\theta)xJ_\nu'([\lambda_n+\theta]x)}{(\lambda_n+\theta)^2}-\frac{J_\nu([\lambda_n+\theta]x)}{(\lambda_n+\theta)^2}\right],$$

and by the asymptotic expansion for Bessel's functions the expression following the sign of equality does not exceed a constant multiple of $1/\lambda_n^{3/2}$ for $0<x\leqq 1$, the constant multiplier being independent of $n$, $x$ and $u$.

**3. The measure of Gibbs's phenomenon.** We prove the following theorem:

THEOREM. *For functions of the class defined in* §1 *and their corresponding developments in Bessel's functions of order* $\nu>0$, *the height of the first wave in the approximation curve,* $y=S_n(x)$, *will for sufficiently large values of* $n$ *be as near as we choose to the value**

$$\nu f(0)\int_0^{\gamma_1}\frac{J_\nu(t)}{t}\,dt, \tag{5}$$

*where* $\gamma_1$ *is the first positive root of the equation* $J_\nu(t)=0$.

As pointed out in §1, the developments in Bessel's functions of a function of the class considered will differ from a series whose general term is

$$\nu\pi f(0)\frac{J_\nu(\lambda_n x)}{\lambda_n} \tag{6}$$

by a series which is uniformly convergent in the neighborhood of the origin. Furthermore the latter series defines in this neighborhood a continuous function which vanishes at the origin. Therefore, for the purpose of studying Gibbs's phenomenon at the origin, the developments in Bessel's functions may be replaced† by the series whose general term is (6). It follows from Theorem I of Transactions I that this latter series may be replaced by the series whose general term is

$$\nu(\lambda_{n+1}-\lambda_n)\frac{J_\nu(\lambda_n x)}{\lambda_n}f(0). \tag{7}$$

From the lemma of §2 and the fact that the $a_n$ there defined is for each $n$ a continuous function vanishing at the origin, it follows that if we first choose $x$ in a sufficiently small interval extending to the right of the origin,

* If the function $f(x)$ has an artificial singularity at the origin, we remove it by defining $f(0)=f(+0)$. In order to infer from (5) that there is actually a Gibbs's phenomenon, we need to be certain that the value of (5) exceeds $f(0)$. That this is the case may be inferred as follows. The sequence

$$\left|\int_{\gamma_n}^{\gamma_{n+1}}[J_\nu(t)/t]\,dt\right|,$$

where the $\gamma_n$'s are the successive positive roots of $J_\nu(t)=0$, is steadily decreasing for increasing $n$ (see Cooke, loc. cit., p. 178). Hence the value of $\int_0^{\gamma_1}[J_\nu(t)/t]dt$ exceeds the value, $1/\nu$, of $\int_0^{\infty}[J_\nu(t)/t]dt$.

† See Bôcher, Annals of Mathematics, (2), vol. 7 (1906), p. 130.

the sum of $n$ terms of the series whose general term is (7) can by a proper choice of $n$ for an $x$ in this interval be made as near as we please to

$$(8) \qquad \nu f(0)\int_0^{\lambda_{n+1}} \frac{J_\nu(xu)}{u}\,du = \nu f(0)\int_0^{\lambda_{n+1}x} \frac{J_\nu(t)}{t}\,dt.$$

Our theorem follows from the fact that the maximum value of the integral $\int_0^y [J_\nu(t)/t]dt$ for values of $y \geqq 0$ is $\int_0^{\gamma_1} [J_\nu(t)/t]dt$ (see first footnote of this section).

**4. Extension to a more general class of functions and to other points of the interval (0, 1).** If we add to the conditions of Lemma 1 of Transactions III the requirement that in an interval including the origin $f(x)$ has a first derivative that is integrable (Lebesgue) and approaches a finite limit as $x$ approaches $+0$, it may be shown by means of formula (15) of Transactions III that for $\nu=0$ or $f(0)=0$ the development in Bessel's functions of $f(x)$ converges uniformly in the neighborhood of the origin.* For $\nu>0$, $f(0)\neq 0$, the development is no longer uniformly convergent on account of the non-vanishing of the fourth term on the right hand side of the formula in question, and since this term is identical with the third term of (1), Gibbs's phenomenon in the neighborhood of the origin is seen to be of the same nature for the more general class of functions considered here.

By combining (1), or (15) of Transactions III, with the asymptotic expansion of the Bessel's function $J_\nu(\lambda_n x)$, it is readily shown that Gibbs's phenomenon at interior points of the interval (0, 1) is identical with the Gibbs's phenomenon of Fourier's cosine development of the same function in the interval (0, 1), the interval of periodicity of the function being of length 2. This will also be the case at the end point $x=1$, if the $l$ of equation (1) of Transactions I is zero; this of course implies that there will be no Gibbs's phenomenon if $f(1)=0$, since the function $f(x)$, defined for all values of $x$ by the periodic property, is then continuous at $x=1$. Also, if $l\neq 0$, there will be no Gibbs's phenomenon, since in this case the Bessel's development is uniformly convergent in the neighborhood of $x=1$, as was shown in Theorem I of Transactions II.

**5. Extension to series summed by Rieszian or Cesàro means.†** We find it more convenient to use in our discussion a Rieszian mean equivalent to the Cesàro mean of order $k(0<k\leqq 1)$, and to consider therefore sums of the type

* This result was contained in a paper presented to the Society in 1917. See Bulletin of the American Mathematical Society, vol. 24 (1918), p. 282.

† The only previous results in this connection are due to Wilton, who shows that in case the index $\nu$ lies in the interval $(0,1/2)$ the $(C1)$ sums do not exhibit Gibbs's phenomenon. (Loc.cit., p. 152.)

$$\sum_{m=1}^{m=n-1}\left(1-\frac{\lambda_m}{\lambda_n}\right)^k u_m(x).$$

On account of the relationship between the two means in question, the results obtained hold equally well for either of them.

We begin by proving a lemma.

LEMMA. *The sum from $m=1$ to $m=n-1$ of terms of the form*

$$(9)\qquad a_m(x)=\int_{\lambda_m}^{\lambda_{m+1}}\left[\left(1-\frac{t}{\lambda_n}\right)^k\frac{J_\nu(xt)}{t}-\left(1-\frac{\lambda_m}{\lambda_n}\right)^k\frac{J_\nu(\lambda_m x)}{\lambda_m}\right]dt$$

*approaches a limit uniformly in the interval $(0\leqq x\leqq 1)$ as $n$ becomes infinite.*

By the law of the mean the expression in the integrand of the above integral may be written in the form

$$(t-\lambda_m)\frac{\dfrac{-kv}{\lambda_n}\left(1-\dfrac{v}{\lambda_n}\right)^{k-1}J_\nu(xv)+vx\left(1-\dfrac{v}{\lambda_n}\right)^k J_\nu'(xv)-\left(1-\dfrac{v}{\lambda_n}\right)^k J_\nu(xv)}{v^2}$$

$$(\lambda_m<v<t<\lambda_{m+1}).$$

For $1\leqq m\leqq n-2$ it follows from the asymptotic expansion for Bessel's functions that the above expression is less in absolute value than a constant multiple of $1/\lambda_m^{3/2}$ for $0<x\leqq 1$, the constant multiple being independent of $m$, $x$, and $t$. For $m=n-1$ it is readily seen that if $H$ represents the maximum absolute value of $J_\nu(u)$ for $u>0$, the term $a_m(x)$ is less in absolute value than $2H(\lambda_n-\lambda_{n-1})/\lambda_{n-1}$. Combining these results and the fact that $a_m(0)=0(m=1, 2, 3, \cdots)$ with the inequalities in (23) of Transactions I, we readily infer that for any positive $\epsilon$ we can so choose $q_1$ that $\left|\sum_{m=p}^{m=n-1}a_m(x)\right|<\frac{1}{3}\epsilon$ $(0\leqq x\leqq 1)$ for any $p\geqq q_1$.

By a similar application of the law of the mean to the integrand of

$$b_m(x)=\int_{\lambda_m}^{\lambda_{m+1}}\left[\frac{J_\nu(xt)}{t}-\frac{J_\nu(\lambda_m x)}{\lambda_m}\right]dt$$

and the use of the asymptotic expansion of $J_\nu(u)$, we readily infer that $b_m(x)$ is the general term of a series uniformly convergent in $(0<x\leqq 1)$ and hence, since $b_m(0)=0$, in $(0\leqq x\leqq 1)$. Therefore we may choose $q_2$ sufficiently large to make $\left|\sum_{m=p}^{\infty}b_m(x)\right|<\frac{1}{3}\epsilon$ in $(0\leqq x\leqq 1)$ for $p\geqq q_2$.

Taking $q$ as the greater of $q_1$ and $q_2$ and holding $q$ fixed, we can then choose a $p>q+1$ so that $\sum_{m=1}^{m=q-1}a_m(x)$ differs by less than $\frac{1}{3}\epsilon$ from $\sum_{m=1}^{m=q-1}b_m(x)$ for all $n\geqq p$. Thus our lemma is established.

For the class of functions considered in the previous section it was the

presence of the term (6) in the general term of the development in Bessel's functions that caused the failure of uniform convergence and hence the appearance of Gibbs's phenomenon in the case $\nu>0, f(0)\neq 0$. Consequently, if we sum the development of a function of the same class by Cesàro means of positive order, the presence or absence of Gibbs's phenomenon and its amount will be determined by using the same method of summation on the series whose general term is (6). Since the term (6) differs from the term

$$\nu f(0)(\lambda_{n+1}-\lambda_n)\frac{J_\nu(\lambda_n x)}{\lambda_n} \tag{10}$$

by the general term of a series that is absolutely and uniformly convergent in the interval $(0\leqq x\leqq 1)$, it is evident that we need only to consider the nature of Gibbs's phenomenon in the following sum:

$$\sum_{m=1}^{m=n-1}\left(1-\frac{\lambda_m}{\lambda_n}\right)^k \nu f(0)(\lambda_{m+1}-\lambda_m)\frac{J_\nu(\lambda_m x)}{\lambda_m}\cdot \tag{11}$$

But since the limit, $\sum_{m=1}^{\infty} b_m(x)$, approached by $\sum_{m=1}^{m=n-1} a_m(x)$ as $n$ becomes infinite, is continuous for $x\geqq 0$ and equal to zero for $x=0$, it follows from our lemma that the sum (11) will for values of $x$ sufficiently small be as near as we choose to

$$\nu f(0)\int_0^{\lambda_n}\left(1-\frac{t}{\lambda_n}\right)^k \frac{J_\nu(xt)}{t}\,dt = \nu f(0)\int_0^{\lambda_n x}\left(1-\frac{u}{\lambda_n x}\right)^k \frac{J_\nu(u)}{u}\,du,$$

when $n$ is chosen sufficiently large.

Hence the presence or absence of Gibbs's phenomenon in the case of summation by Rieszian or Cesàro means of order $k>0$ will depend on the behavior of the function $\phi_k(\lambda)$ defined by

$$\phi_k(\lambda) = \int_0^{\lambda}\left(1-\frac{u}{\lambda}\right)^k \frac{J_\nu(u)}{u}du.$$

If for all positive values of $\lambda$, $\phi_k(\lambda)$ remains less than or equal to its limiting value as $\lambda$ becomes infinite, $1/\nu$, there will be no Gibbs's phenomenon. If for certain positive values of $\lambda$ it rises above this limit, there will be a Gibbs's phenomenon and the measure of this phenomenon will be determined by the maximum value of $\phi_k(\lambda)$.

Consider first the case $k=1$. We have

$$\phi_1(x) = \int_0^x\left(1-\frac{u}{x}\right)\frac{J_\nu(u)}{u}du = \int_0^1 (1-t)\,\frac{J_\nu(tx)}{t}dt,$$

$$\phi_1'(x) = \int_0^1 (1-t)J_\nu'(tx)dt = \left[\frac{(1-t)J_\nu(tx)}{x}\right]_0^1 + \frac{1}{x}\int_0^1 J_\nu(tx)dt$$

$$= \frac{1}{x}\int_0^1 J_\nu(tx)dt = \frac{1}{x^2}\int_0^x J_\nu(u)du.$$

Since $\int_0^\infty J_\nu(u)du$ is positive for all positive values of $x$ (see reference in first footnote of §3), it is apparent that $\phi_1(x)$ is a steadily increasing function of $x$ for such values and that therefore *there is no Gibbs's phenomenon when we sum the developments in Bessel's functions by Rieszian or Cesàro means of order* 1.

Consider now values of $k$ in the interval $0<k<1$. The function $\phi_k(x)$ is a continuous function of $x$ and $k$ for $x>0$, $k\geqq 0$. Hence, for sufficiently small values of $k$ it will oscillate about its limiting value, $1/\nu$, just as $\phi_0(x)$ does. Therefore *we have the Gibbs's phenomenon for values of k near zero.*

That this is not the case for values of $k$ near 1 we prove as follows. We have

$$\phi_k'(x) = \int_0^1 (1-t)^k J_\nu'(tx)dt = \left[\frac{(1-t)^k J_\nu(tx)}{x}\right]_0^1 + \frac{k}{x}\int_0^1 (1-t)^{k-1}J_\nu(tx)dt$$

(12)

$$= \frac{k}{x^2}\int_0^x \left(1-\frac{u}{x}\right)^{k-1} J_\nu(u)du.$$

It can readily be shown that for values of $k>\frac{1}{2}$ the integral on the right hand side of (21) approaches $\int_0^\infty J_\nu(u)du = 1$ as $x$ becomes infinite.

We divide the interval of integration into four parts $(0, l)$, $(l, x/2)$, $(x/2, x-1)$, $(x-1, x)$ and write

$$\int_0^x \left(1-\frac{u}{x}\right)^{k-1} J_\nu(u)du = \int_0^l + \int_l^{x/2} + \int_{x/2}^{x-1} + \int_{x-1}^x = I_1 + I_2 + I_3 + I_4.$$

Applying the second law of the mean to $I_2$ and $I_3$ and the first law of the mean of $I_4$, we obtain

$$I_2 = 2^{1-k}\int_y^{x/2} J_\nu(u)du \quad (y > l), \qquad I_3 = x^{1-k}\int_\xi^{x-1} J_\nu(u)du \qquad (\xi > \tfrac{1}{2}x),$$

$$I_4 = \frac{J_\nu(x-\theta)x^{1-k}}{k} \qquad (0 < \theta < 1).$$

$I_2$ obviously approaches zero as $l$ becomes infinite, and it follows from the asymptotic expansion of $J_\nu(u)$ that, if $k>\frac{1}{2}$, $I_3$ and $I_4$ approach zero as

$x$ becomes infinite. If we now allow $x$ and $l$ to become infinite in such a manner that $x/l$ becomes infinite, $I_1$ approaches $\int_0^\infty J_\nu(u)du$, and therefore we have

$$\lim_{x\to\infty}\int_0^x\left(1-\frac{u}{x}\right)^{k-1}J_\nu(u)du=\int_0^\infty J_\nu(u)du=1 \qquad (k>\tfrac{1}{2}).$$

It follows from the above relationship that we can find an $X$ such that $\int_0^x(1-u/x)^{k-1}J_\nu(u)du$ remains positive for $x\geqq X$ when $k>\frac{1}{2}$. Moreover, for values of $x<X$, we can choose $k$ sufficiently near 1 to make $\int_0^x(1-u/x)^{k-1}\cdot J_\nu(u)du$ differ from $\int_0^x J_\nu(u)du$ by as small a quantity as we please throughout the interval $0<x<X$. But $\int_0^x J_\nu(u)du$ is positive for all $x>0$ and approaches 1 as $x$ becomes infinite. Hence it has a lower limit $\rho>0$ for all $x\geqq\delta>0$. Consequently, for $k$ near enough to 1, $\int_0^x(1-u/x)^{k-1}J_\nu(u)du$ will also remain positive for $x\geqq\delta$, where $\delta$ is any given positive quantity. But for a $\delta$ less than the first positive root of $J_\nu(u)=0$, this integral is obviously positive for $0<x<\delta$ and any $k>0$. Thus *for values of $k$ sufficiently near 1 the function $\phi_k(x)$ steadily increases with $x$ and there is no Gibbs's phenomenon.*

We wish finally to show that there is a value $r$ between 0 and 1 such that Gibbs's phenomenon occurs for $k<r$ and does not occur for $k\geqq r$. Let us set

$$M_k(x,n)=\sum_{m=1}^{m=n-1}\left(1-\frac{\lambda_m}{\lambda_n}\right)^k(\lambda_{m+1}-\lambda_m)\frac{J_\nu(\lambda_m x)}{\lambda_m}.$$

Combining the lemma of this section with the fact that the $a_m(x)$ defined by (9) is zero for $x=0$, we obtain

$$h(k)=\limsup_{\substack{x\to 0\\ n\to\infty}}M_k(x,n)=\limsup_{x>0}\phi_k(x).$$

But if we write $M_{k+\delta}(x,n)$ in the form

$$\sum_{m=1}^{m=n-1}\left(1-\frac{\lambda_m}{\lambda_n}\right)^\delta\left(1-\frac{\lambda_m}{\lambda_n}\right)^k(\lambda_{m+1}-\lambda_m)\frac{J_\nu(\lambda_m x)}{\lambda_m},$$

it follows from Abel's lemma that $h(k+\delta)\leqq h(k)$. Hence $h(k)$ is a monotonic decreasing function of $k$ in the interval $0<k<1$. Moreover, since $\phi_k(x)$ is a continuous function of $x$ and $k$ for $x>0$, $k\geqq 0$, $h(k)$ is a continuous function of $k$ for $k\geqq 0$. But $h(k)=1/\nu$ for values of $k$ near 1 and $h(0)>1/\nu$. Consequently there is a value $r$ of $k$ between 0 and 1 such that $h(k)=1/\nu (r\leqq k\leqq 1)$ and $h(k)>1/\nu$ $(0<k<r)$. Thus *Gibbs's phenomenon will occur for $k<r$ and will not occur for $k\geqq r$.*

UNIVERSITY OF CINCINNATI,
CINCINNATI, OHIO

# SURFACE TRANSFORMATIONS*

BY

F. R. BAMFORTH†

In a paper written by Birkhoff‡ on the Problem of Three Bodies, he reduced the study of a certain phase of this celebrated problem to the study of a surface transformation which was the product of two surface transformations each of which represented the reflection, in a curve lying in a surface, of the surface into itself. It is the purpose of this paper to make a start on the study of such surface transformations in general.

In the first section are studied surface transformations of period two, and it is shown that when a suitable coördinate system has been chosen, any such transformation may be represented by one and but one of the following systems of equations:

$$(1)\qquad \begin{array}{lll} u_1 = u, & u_1 = -u, & u_1 = u, \\ v_1 = v; & v_1 = -v; & v_1 = -v; \end{array}$$

in which the point $(u_1, v_1)$ denotes the transform of the point $(u, v)$.

In the remaining sections, surface transformations are discussed which are the product of two transformations, each of which, when a suitable coördinate system has been selected, can be represented by equations of the third type in (1).

## I. Transformations of period two

This section will be concerned with real, one-to-one, analytic transformations of a surface $Z$ into itself, of such a character that after the same transformation has been performed twice in succession all the points of the surface are in the same positions with respect to fixed axes as they were before. These transformations are called transformations of period two, and we shall study the movement under such transformations of the points of the surface $Z$ in a neighborhood of a point which is invariant.

In a neighborhood of such an invariant point the equations representing the transformation may be written in the form

$$(2)\qquad u_1 = au + bv + \cdots, \qquad v_1 = cu + dv + \cdots, \qquad ad - bc \neq 0,$$

* Presented to the Society, December 27, 1929; received by the editors February 28, 1930.

† This paper was partially prepared while the author was a National Research Fellow.

‡ G. D. Birkhoff, Rendiconti del Circolo Matematico di Palermo, vol. 39.

after a proper choice of coördinate system has been made. For this system of coördinates the origin is the invariant point and the series on the right hand sides of the equations (2) are convergent for $u$ and $v$ sufficiently small in absolute value. By means of a real, linear change of variables, the equations (2) of the transformation are equivalent to equations which will have one of the following forms:

$$u_1 = \alpha u + \cdots, \qquad v_1 = \alpha v + \cdots, \ \alpha \neq 0; \tag{3}$$

$$u_1 = v + \cdots, \qquad v_1 = \beta u + \delta v + \cdots, \ \beta \neq 0, \tag{4}$$

where the dots indicate terms of degree greater than one.

Let us now discuss transformations which may be represented by equations of the form (3). If we denote the transform of $(u_1, v_1)$ by $(u_2, v_2)$, from the fact that the transformation is of period two, it follows that

$$u = u_2 = \alpha u_1 + \cdots = \alpha^2 u + \cdots, \qquad v = v_2 = \alpha v_1 + \cdots = \alpha^2 v + \cdots,$$

whence $\alpha = \pm 1$. Suppose first that $\alpha = 1$. The transformation may be represented by

$$u_1 = u + a_{20}u^2 + a_{11}uv + a_{02}v^2 + \cdots,$$
$$v_1 = v + b_{20}u^2 + b_{11}uv + b_{02}v^2 + \cdots.$$

Since the transformation is of period two

$$\begin{aligned} u = u_2 &= u_1 + a_{20}u_1^2 + a_{11}u_1v_1 + a_{02}v_1^2 + \cdots \\ &= u + 2a_{20}u^2 + 2a_{11}uv + 2a_{02}v^2 + \cdots. \end{aligned}$$

From this identity it follows that all the $a_{ij}$ are zero, and from the corresponding one in $v$ it follows that all the $b_{ij}$ are also zero. Hence, when a transformation of period two can be represented by equations of the form (3) with $\alpha = 1$, it is the identity transformation.

Transformations represented by equations of the form (3) with $\alpha = -1$ will be called "rotation transformations" on account of the fact that the simplest example of equations of this form, i.e., $u_1 = -u$, $v_1 = -v$, represents a rotation through 180° when the coördinate system is rectangular. We shall now reduce the equations (3) of a rotation transformation to a simpler form. To this end let us write these equations in the form

$$u_1 = -u + U(u,v), \qquad V_1 = -v + V(u,v), \tag{5}$$

where $U(u,v)$ and $V(u,v)$ have no terms of first degree in $u$ and $v$. On account of the periodicity of the transformation,

$$u = u_2 = -u_1 + U(u_1,v_1) = u - U(u,v) + U(u_1,v_1),$$
$$v = v_2 = -v_1 + V(u_1,v_1) = v - V(u,v) + V(u_1,v_1),$$

whence

$$U(u,v) = U(u_1,v_1), \quad V(u,v) = V(u_1,v_1).$$

Hence the equations (5) are equivalent to the equations

$$u_1 - \tfrac{1}{2}U(u_1,v_1) = -u + \tfrac{1}{2}U(u,v),$$
$$v_1 - \tfrac{1}{2}V(u_1,v_1) = -v + \tfrac{1}{2}V(u,v).$$

Under the change of variables

$$u^* = u - \tfrac{1}{2}U(u,v), \qquad v^* = v - \tfrac{1}{2}V(u,v),$$

these equations are evidently transformed into

$$u_1^* = -u^*, \quad v_1^* = -v^*,$$

from which form the structure of the transformation is evident.

If a transformation of period two is represented by equations of the form (4),

$$u = u_2 = v_1 + \cdots = \beta u + \delta v + \cdots,$$
$$v = v_2 = \beta u_1 + \delta v_1 + \cdots = \beta v + \delta\beta u + \delta^2 v + \cdots,$$

whence $\beta = 1$ and $\delta = 0$. The equations representing such a transformation, by means of a suitable change of coördinate system, may be of either of the two forms

$$u_1 = v + \cdots, \quad v_1 = u + \cdots, \tag{6}$$

or

$$u_1 = u + \cdots, \quad v_1 = -v + \cdots. \tag{7}$$

The equivalence is readily seen if the coördinates are considered as rectangular coördinates in the plane and a rotation of axes through an angle of 45° is performed. Since the simplest example of equations of the form (7) representing such a transformation of period two, i.e., $u_1 = u$, $v_1 = -v$, represents a reflection in the $u$ axis, we shall call any transformation of period two which can be represented by equations of the form (6) or (7), a "reflection transformation."

Let us write the equations (7) in the form

$$u_1 = u + U(u,v), \qquad v_1 = -v + V(u,v). \tag{8}$$

Here, on account of the periodicity of the transformation,

$$U(u,v) = -U(u_1,v_1), \quad V(u,v) = V(u_1,v_1).$$

Hence the equations (8) are equivalent to the equations

$$u_1 + \tfrac{1}{2}U(u_1,v_1) = u + \tfrac{1}{2}U(u,v),$$
$$v_1 - \tfrac{1}{2}V(u_1,v_1) = -v + \tfrac{1}{2}V(u,v),$$

from which it follows immediately that this transformation may be represented by the equations

$$u_1^* = u^*, \quad v_1^* = -v^*, \tag{9}$$

where

$$u^* = u + \tfrac{1}{2}U(u,v), \qquad v^* = v - \tfrac{1}{2}V(u,v).$$

Here, as in the case of rotation transformations, the form (9) tells us everything that we wish to know about the transformation.

When a transformation of the reflection type is represented in a neighborhood of an invariant point by equations of the form (9), we shall say that the $u^*v^*$ system of coördinates is a "normal system" for the transformation. Now it is evident that there are infinitely many normal coördinate systems for any one such transformation. In fact, if we make a reversible change of coördinates defined by the equations

$$u^* = f(w,z), \qquad v^* = g(w,z), \tag{10}$$

a sufficient condition that the new system of coördinates be a normal system for this transformation is that

$$f(w,-z) = f(w,z), \quad g(w,-z) = -g(w,z). \tag{11}$$

This is a consequence of the fact that if (11) be true, $w_1 = w$, $z_1 = -z$ is a solution of the pair of equations

$$f(w_1,z_1) = f(w,z), \qquad g(w_1,z_1) = -g(w,z).$$

By way of summarizing this first section, we may say that *every transformation of period two, in a neighborhood of an invariant point, after a suitable choice of coördinate system has been made, may be represented by one and but one of the following systems of equations*:

$$u_1 = u, \qquad u_1 = -u, \qquad u_1 = u,$$
$$v_1 = v; \qquad v_1 = -v; \qquad v_1 = -v.$$

## II. Product transformations

In Birkhoff's paper on the Problem of Three Bodies, a transformation of a surface into itself arose which was the product of two transformations each of which was of the reflection type. The object of this and the succeeding sections of this paper is to study such transformations in the neighborhood

of a point which is invariant under each of the component transformations of period two.

In the notation we shall use, the script letters $\mathfrak{R}$ and $\mathfrak{S}$ will stand for transformations of the reflection type of a surface $Z$ into itself, and the script letter $\mathfrak{T}$ will stand for the product transformation $\mathfrak{T}=\mathfrak{R}\mathfrak{S}$. In a neighborhood of a point which is invariant under the transformation $\mathfrak{S}$, by choosing a normal system of coördinates for this transformation, it may be represented by the equations

$$S_0: \quad y_1 = y, \quad x_1 = -x. \tag{12}$$

Here, $S_0$ is interpreted as symbolical of the equations (12). If we make a reversible change of coördinates defined by

$$V: \quad y = f(w,z), \quad x = g(w,z), \tag{13}$$

the equations (12) are equivalent to the equations

$$f(w_1,z_1) = f(w,z), \quad g(w_1,z_1) = -g(w,z). \tag{14}$$

Since we have denoted the equations (13) symbolically by $V$, let us denote the equations obtained by solving them for $w$ and $z$ in terms of $y$ and $x$ by $V^{-1}$, i.e.,

$$V^{-1}: \quad w = f_0(y,x), \quad z = g_0(y,x).$$

On solving (14) for $w_1$ and $z_1$ we obtain

$$w_1 = f_0[f(w,z), -g(w,z)],$$
$$z_1 = g_0[f(w,z), -g(w,z)].$$

These equations may be represented symbolically by $V^{-1}S_0V$ when this symbol is interpreted according to the law which will now be stated. If $A$, $B$ and $C$ denote the systems of equations

$$A: \begin{matrix}\alpha = f_1(\gamma,\delta)\\ \beta = g_1(\gamma,\delta)\end{matrix}, \quad B: \begin{matrix}\gamma = f_2(\xi,\eta)\\ \delta = g_2(\xi,\eta)\end{matrix}, \quad C: \begin{matrix}\xi = f_3(\zeta,\theta)\\ \eta = g_3(\zeta,\theta)\end{matrix},$$

then $AB$ and $ABC$ denote, respectively, the systems of equations

$$AB: \begin{matrix}\alpha = f_1[f_2(\xi,\eta), g_2(\xi,\eta)],\\ \beta = g_1[f_2(\xi,\eta), g_2(\xi,\eta)];\end{matrix}$$

and

$$ABC: \begin{matrix}\alpha = f_1\{f_2[f_3(\zeta,\theta), g_3(\zeta,\theta)], g_2[f_3(\zeta,\theta), g_3(\zeta,\theta)]\},\\ \beta = g_1\{f_2[f_3(\zeta,\theta), g_3(\zeta,\theta)], g_2[f_3(\zeta,\theta), g_3(\zeta,\theta)]\}.\end{matrix}$$

Thus we have the result that if a transformation $\mathfrak{S}$ of a surface $Z$ into

itself be denoted by the equations $S_0$ of (12), and if a change of coördinate systems denoted by the equations $V$ of (13) takes place, the transformation $\mathfrak{S}$ is denoted in the new system of coördinates by the equations $V^{-1}S_0V$.

Let us suppose now that the transformation $\mathfrak{R}$ is represented by the system of equations

$$R: \quad u_1 = u, \quad v_1 = -v, \tag{15}$$

and let the $yx$ and the $uv$ systems of coördinates be related by means of the equations

$$U: \quad y = F(u,v), \quad x = G(u,v), \tag{16}$$

where $F$ and $G$ are analytic functions of their arguments in a neighborhood of the point $(u, v) = (0, 0)$, the jacobian

$$\left| \begin{matrix} \partial F/\partial u & \partial F/\partial v \\ \partial G/\partial u & \partial G/\partial v \end{matrix} \right|_{(u,v)=(0,0)} \tag{17}$$

is greater than zero, and $F(0, 0) = G(0, 0) = 0$. The condition that the jacobian (17) be greater than zero is not so strong as one might imagine. If we have a correspondence between the two systems of coördinates which fulfills the other conditions named above but which is such that the jacobian (17) is negative, we may change our $yx$ coördinate system to another by means of the equations

$$y = w, \quad x = -z.$$

The $wz$ system of coördinates will be normal for $\mathfrak{S}$ and the analogue of the negative jacobian will be positive.

Since the $yx$ normal coördinate system can be replaced by another by means of the equations (10), where $f$ and $g$ satisfy the relations (11), it it evident that we may assume that the determinant of the coefficients of the linear terms of the functions $F$ and $G$ in (16) has the value 1, and that, if the linear term in $u$ (or the linear term in $v$) is lacking either in $F$ or in $G$, the coefficient of the linear term in $v$ (or the linear term in $u$) is unity in absolute value. Thus there exist what may be termed certain canonical forms for the equations (16) which are characterized by the linear terms of $F$ and $G$. These may be listed as:

$$\text{I(A)}: \quad \begin{aligned} y &= u + \cdots, \\ x &= v + \cdots; \end{aligned} \tag{18}$$

$$\text{I(B)}: \quad \begin{aligned} y &= u + av + \cdots, \quad a \neq 0, \\ x &= v + \cdots; \end{aligned}$$

$$\text{II(A):}\quad \begin{aligned} y &= v + \cdots, \\ x &= -u + \cdots; \end{aligned}$$

$$\text{II(B):}\quad \begin{aligned} y &= au + v + \cdots, \quad a \neq 0, \\ x &= -u + \cdots; \end{aligned}$$

$$\text{III:}\quad \begin{aligned} y &= u + \cdots, \\ x &= au + v + \cdots, \quad a \neq 0; \end{aligned}$$

$$\text{IV:}\quad \begin{aligned} y &= v + \cdots, \\ x &= -u - av + \cdots, \quad a \neq 0; \end{aligned}$$

$$\text{V:}\quad \begin{aligned} y &= a_{10}u + a_{01}v + \cdots, \\ x &= b_{10}u + b_{01}v + \cdots, \quad a_{10}a_{01}b_{10}b_{01} > 0; \end{aligned}$$

$$\text{VI:}\quad \begin{aligned} y &= a_{10}u + a_{01}v + \cdots, \\ x &= b_{10}u + b_{01}v + \cdots, \quad a_{10}a_{01}b_{10}b_{01} < 0. \end{aligned}$$

The equations I(B), II(B), III and IV may be assumed to be in a still simpler form. Consider, for example, the equations I(B), and make use of the change of coördinates

$$\begin{aligned} y &= y^*, & u &= u^*, \\ x &= x^*/(2a); & v &= v^*/(2a). \end{aligned}$$

Evidently the transformed equations of I(B) have the form

$$\begin{aligned} y^* &= u^* + \tfrac{1}{2}v^* + \cdots, \\ x^* &= v^* + \cdots. \end{aligned}$$

Since analogous transformations may be employed in the other cases, it is evident that *we may assume that wherever "a" appears in the equations* (18) *we may write* $\frac{1}{2}$. This choice of constant "$a$" gives a particularly simple form to the canonical forms which will be given later of the equations representing transformations $\mathfrak{T}$.

Let us now consider the transformation $\mathfrak{T}$ of the surface $Z$ into itself which is the product of the transformations $\mathfrak{R}$ and $\mathfrak{S}$, i.e., $\mathfrak{T}=\mathfrak{R}\mathfrak{S}$, which is to mean that the transformation $\mathfrak{S}$ is followed by the transformation $\mathfrak{R}$. We shall first discuss one of the properties of this transformation $\mathfrak{T}$ which can be readily proved out of the fact that each of the transformations $\mathfrak{R}$ and $\mathfrak{S}$ is of period two. Let the symbol $S$ represent the equations of the transformation $\mathfrak{S}$ in the $uv$ system of coördinates which is normal for the transformation $\mathfrak{R}$ represented by the equations (15). On account of the equations (16) we see that symbolically $S=U^{-1}S_0U$, where $S_0$ is given in (12). If $T$

represents the equations of the transformation $\mathfrak{T}$ in the $uv$ system of coördinates, evidently $T=RS$. If we denote the identity transformation by $I$, we have symbolically

$$SS = I,\ S = S^{-1},\ R = R^{-1},\ T = RS,\ TSR = RSSR = I,$$
$$T^{-1} = SR = SRSS^{-1} = RRSR = STS^{-1} = R^{-1}TR.$$

It is this last equation, $T^{-1}=R^{-1}TR$, which we wish to examine more closely. Since a change of coördinate systems denoted by $V$ changes the equations $T$ into $V^{-1}TV$ we see that the inverse transformation of $\mathfrak{T}$ is represented in the $uv$ system of coördinates by the same equations as the transformation $\mathfrak{T}$ after the change of coördinate systems denoted by $R$ has taken place. Thus, if the equations $T$ are

$$T: \quad \begin{aligned} u_1 &= a_1u + a_2v + \cdots, \\ v_1 &= b_1u + b_2v + \cdots, \end{aligned} \tag{19}$$

and if we perform a change of coördinates denoted by

$$R: \quad u = s,\ v = -t, \tag{20}$$

obtaining thereby

$$R^{-1}TR: \quad \begin{aligned} s_1 &= a_1s - a_2t + \cdots, \\ t_1 &= -b_1s + b_2t + \cdots, \end{aligned} \tag{21}$$

in the $uv$ system of coördinates $\mathfrak{T}^{-1}$ is given by

$$T^{-1}: \quad \begin{aligned} u &= a_1u_1 - a_2v_1 + \cdots, \\ v &= -b_1u_1 + b_2v_1 + \cdots. \end{aligned} \tag{22}$$

Thus, observing the movement of the points of the surface $Z$ under the transformation $\mathfrak{T}^{-1}$ is equivalent to observing the movement of the points of the reflection of the surface $Z$ under $\mathfrak{T}$ in a mirror which is plane for the $uv$ system of coördinates. We shall make use of this fact later on in the discussion.

Let us now return to a further consideration of the normal forms (18) of the equations of the change of coördinate systems. The transformation $\mathfrak{S}$ in the $uv$ system is given by

$$F(u_1,v_1) = F(u,v), \qquad G(u_1,v_1) = -G(u,v). \tag{23}$$

Hence the transformation $\mathfrak{T}$ is given by

$$F(u_1, -v_1) = F(u,v), \qquad G(u_1, -v_1) = -G(u,v). \tag{24}$$

The normal forms for the equations $T$ which correspond to the normal forms

(18) are the following:

$$
\begin{aligned}
&\text{I(A):} && u_1 = u + \cdots, \\
& && v_1 = v + \cdots; \\
&\text{I(B):} && u_1 = u + v + \cdots, \\
& && v_1 = v + \cdots; \\
&\text{II(A):} && u_1 = -u + \cdots, \\
& && v_1 = -v + \cdots; \\
&\text{II(B):} && u_1 = -u + \cdots, \\
& && v_1 = -u - v + \cdots; \\
&\text{III:} && u_1 = u + \cdots, \\
& && v_1 = u + v + \cdots; \\
&\text{IV:} && u_1 = -u - v + \cdots, \\
& && v_1 = -v + \cdots; \\
&\text{V:} && u_1 = \alpha u + \beta v + \cdots, \\
& && v_1 = \gamma u + \alpha v + \cdots, \ \beta\gamma > 0, \ \alpha \neq 0; \\
&\text{VI:} && u_1 = \alpha u + \beta v + \cdots, \\
& && v_1 = \gamma u + \alpha v + \cdots, \ \beta\gamma < 0;
\end{aligned}
\tag{25}
$$

where $\alpha = a_{10}b_{01} + a_{01}b_{10}$, $\beta = 2a_{01}b_{01}$, $\gamma = 2a_{10}b_{10}$, $\alpha^2 - \beta\gamma = 1$. Evidently, each of the type sets of equation in (25) is but an example of

$$
\begin{aligned}
&T: && u_1 = \alpha u + \beta v + \cdots, \\
& && v_1 = \gamma u + \alpha v + \cdots,
\end{aligned}
\tag{26}
$$

where $\alpha, \beta$ or $\gamma$ may have the value zero.

Thus we have obtained the result that *every transformation which is the product of two transformations of the reflection type, in a neighborhood of a point which is invariant under each of the component transformations, can be represented in one and only one way by equations which are of one of the types displayed in* (25).

## III. Invariant directions and simple examples

Let us now inquire into the possibility of there being directions through the origin which are either invariant under $T$ or are rotated through 180°. From (26) when $\alpha\beta\gamma \neq 0$ we obtain at $(u, v) = (0, 0)$ for $du/dv$ finite and different from $-\alpha/\gamma$

$$\frac{du_1}{dv_1}-\frac{du}{dv}=\frac{\alpha du+\beta dv}{\gamma du+\alpha dv}-\frac{du}{dv}=\frac{\beta-\gamma\left(\frac{du}{dv}\right)^2}{\gamma\frac{du}{dv}+\alpha}.$$

If $du/dv=-\alpha/\gamma$ or is infinite, then $dv/du\neq-\alpha/\beta$, since $\alpha^2-\beta\gamma=1$, and

$$\frac{dv_1}{du_1}-\frac{dv}{du}=\frac{\gamma-\beta\left(\frac{dv}{du}\right)^2}{\alpha+\beta\frac{dv}{du}}.$$

Hence $du/dv$ remains unchanged by $T$ if and only if

$$\beta(dv)^2-\gamma(du)^2=0. \tag{27}$$

Now consider the case where $\alpha=0$. Then $\beta\gamma\neq0$, and

$$\frac{du_1}{dv_1}=\frac{\beta}{\gamma}\frac{dv}{du}$$

when $dv/du$ is finite and

$$\frac{dv_1}{du_1}=\frac{\gamma}{\beta}\frac{du}{dv}$$

when $du/dv$ is finite. Thus the invariant values of $du/dv$ are given by (27). In a similar manner it can readily be shown that the invariant values of $du/dv$ for $\beta=0$ or $\gamma=0$ are also given by (27).

Thus we find that *for all transformations of the types* I(A) *and* II(A), *all slopes through the origin are invariant*; *for all transformations of the types* I(B), II(B), III *and* IV *there are two coincident invariant slopes which are the only ones*; *for all transformations of the type* V *there are two distinct invariant slopes which are the only ones*; *and for transformations of the type* VI *there is no invariant slope.*

It would, perhaps, be useful in clarifying our ideas to give at this point a short discussion of transformations of the different fundamental types in which only linear terms appear on the right hand sides of the equations (25). Each of these transformations is area preserving since the jacobian of the functions on the right hand sides of each pair of these equations is 1.

The transformation $u_1=u$, $v_1=v$, which is of type I(A), is evidently the identity transformation under which every point is an invariant point.

For the transformation

$$u_1=u+v,\quad v_1=v, \tag{28}$$

of type I(B), the $u$ and $y$ axes are coincident while the equation of the $x$-axis in the $uv$-system of coördinates is given by $u+\frac{1}{2}v=0$. The only invariant curve for this transformation which passes through the origin is the $u$-axis, and since $v$ is an invariant function, the points move parallel to the $u$-axis under the transformation $\mathfrak{T}$. Moreover, this invariant curve is made up of invariant points. It is easily verified that this transformation has the origin as a hyperbolic, unstable,† invariant point, since for $v>0$, $u_1-u>0$, and for $v<0$, $u_1-u<0$. In fact, if any neighborhood of the origin be taken, it is evident that the only points which remain in the neighborhood under the indefinite iteration of $T$ and of $T^{-1}$ are those lying on the $u$-axis which are invariant points.

The transformation $u_1=-u$, $v_1=-v$, of type II(A), is a rotation through 180°. Thus every straight line through the origin is invariant and the points on it are reflected in the origin, which is an isolated, hyperbolic, stable, invariant point for $\mathfrak{T}$. The $x$ and the $u$ axes are coincident, as are also the $y$ and the $v$.

For the transformation $u_1=-u$, $v_1=-u-v$, of type II(B), the $y$ and $v$ axes are coincident and the equation of the $x$-axis in the $uv$-system of coördinates is given by $\frac{1}{2}u+v=0$. The only invariant curve through the origin is the $v$-axis on which all the points are reflected in the origin by $\mathfrak{T}$. The origin is a hyperbolic, unstable, invariant point.

For the transformation $u_1=u$, $v_1=u+v$, of type III, the $x$ and $v$ axes are coincident and the equation of the $y$ axis is $\frac{1}{2}u+v=0$. The origin is a hyperbolic, unstable, invariant point, and the only invariant curve through it is the $v$ axis. The discussion of the simple transformation of type IV is obviously similar to that of type III.

For the transformation

$$u_1 = \alpha u + \beta v, \quad v_1 = \gamma u + \alpha v, \quad \alpha > 0, \quad \beta > 0, \quad \gamma > 0, \tag{29}$$

of type V, $\beta v^2-\gamma u^2$ is an invariant function which is zero on the two invariant lines $\beta v^2-\gamma u^2=0$ through the origin, which is therefore a hyperbolic, invariant point. The $u$ axis bisects the angle between these two invariant lines. Let us examine the movement under $\mathfrak{T}$ of the points on these invariant lines. Let $(u, v)$ be any point on the invariant line $u=v(\beta/\gamma)^{1/2}$. Then

$$u_1 = \alpha u + \beta v = (\alpha + (\beta\gamma)^{1/2})u,$$
$$v_1 = \gamma u + \alpha v = ((\beta\gamma)^{1/2} + \alpha)v.$$

† Definitions of certain technical terms used in the discussion of this paper will be found in a paper by G. D. Birkhoff, Acta Mathematica, vol. 43: conservative transformation, p. 2; formal conservative transformation, p. 22; quasi-invariant function, p. 2; stable and unstable invariant points, p. 5; formal invariant curve through origin, p. 26; hypercontinuous invariant curve, p. 67; elliptic and hyperbolic invariant points, p. 26. This paper will be referred to hereafter as B.

Now $(\alpha+(\beta\gamma)^{1/2})\ (\alpha-(\beta\gamma)^{1/2})=1$. Hence $\alpha+(\beta\gamma)^{1/2}$ is positive and either is greater than 1 or less than 1. If $\alpha+(\beta\gamma)^{1/2}>1$, the points on this invariant line will recede from the origin on iteration of $\mathfrak{T}$ and if $\alpha+(\beta\gamma)^{1/2}<1$ the points will approach the origin on iteration of $\mathfrak{T}$. But if $\alpha+(\beta\gamma)^{1/2}<1$, $\alpha-(\beta\gamma)^{1/2}>1$, and hence the points on one of the invariant lines recede from the origin on iteration of $\mathfrak{T}$ and the points on the other invariant line approach the origin on iteration of $\mathfrak{T}$. This shows that the origin is an unstable invariant point.

Suppose that for the transformation represented by (29) we make a change of coördinates which takes these invariant curves into the axes. Then the equations (29) become $\bar{u}_1=b\bar{u}$, $\bar{v}_1=c\bar{v}$, $b>1$, $0<c<1$, which shows that no other invariant curves exist for the transformation (29) than those already mentioned.

The transformation which is the same as (29) with the exception that $\alpha<0$, can be discussed in a manner similar to that in which we have discussed (29). In this transformation the points on one of the invariant lines recede from the origin on iteration of $\mathfrak{T}$ but oscillate about the origin in so doing. By this is meant that the origin always separates a point on this invariant line from its image under $\mathfrak{T}$. The points on the other invariant line approach the origin under iteration of $\mathfrak{T}$ but also oscillate with respect to it.

For the transformation $u_1=\alpha u+\beta v$, $v_1=\gamma u+\alpha v$, $\beta>0$, $\gamma<0$, which is of the type VI, as has already been proved there is no invariant slope through the origin. Hence there is no invariant curve through it and the origin is an elliptic, invariant point. In this case the function $\beta v^2-\gamma u^2$ is an invariant function which shows that the origin is a stable invariant point.

Since all the simple transformations which we have just been discussing preserve areas and hence have invariant integrals, it would seem that it would be advantageous to link that classification which we have given for transformations $\mathfrak{T}=\mathfrak{RS}$ with that given by Birkhoff in his paper in the Acta Mathematica. It will be assumed that the reader has this paper before him; the classification is on page 4. We shall designate the types of problems which we are studying by I(A), I(B), etc., and those which Birkhoff studied by I′, I″, etc., as he did. With this notation in mind, it is evident that we have the following correspondence based upon the coefficients of the linear terms in the equations of the transformations: I(A)$\sim$II″, I(B)$\sim$III′, II(A)$\rightarrow$II‴, II(B)$\sim$III″, III$\sim$III′, IV$\sim$III″, V($\alpha>0$)$\sim$I′, V($\alpha<0$)$\sim$I″, VI($\alpha=0$)$\sim$II‴, VI($\alpha\neq0$)$\sim$II′ or II‴. Reference will be made to this correspondence from time to time. When interpreting this correspondence we should remember the properties of the $u$ and $v$ axes of coördinates as used in this paper, made clear by the equations (15).

## IV. Formal invariant series

The question of the existence of real, formal series which are invariant under the equations representing a transformation $\mathfrak{T}$ is important for at least two reasons. If there is such a series which is convergent, by setting it equal to a constant we obtain a curve which is invariant under $\mathfrak{T}$. Hence we can study the transformation by studying the totality of such invariant curves. On the other hand, if there does not exist any series which is convergent, but there does exist one that is divergent for all sets of values of the variables different from $(0, 0)$, it can be used in many cases to reduce the equations representing the transformation to a simpler form and to prove the existence or non-existence of real invariant curves through the origin which is assumed to be an invariant point of the transformation. Hence we shall search for series which are formally invariant under $T=RS$.

Let us now recall that $T^{-1}=R^{-1}TR=S^{-1}TS$, which means that the equations giving the inverse transformation $\mathfrak{T}^{-1}$ are the same in form as those for the transformation $\mathfrak{T}$ itself after a change of variables has taken place in which the equations of the change of variables are those of the transformation $\mathfrak{R}$ or of the transformation $\mathfrak{S}$. In this connection, let us return to the consideration of the equations (19), (20), (21) and (22), and the discussion given concerning them. Thus we see that, if $H(u, v)$ is a formal invariant series under (19), $H(s,-t)$ is invariant under (21) and hence $H(u,-v)$ is invariant under (22). Thus, if the series $H(u, v)$ is invariant under $T$, the series $H(u,-v)$, $H(u, v)+H(u,-v)$ and $H(u, v)-H(u,-v)$ are too; and every series which is invariant under $T$ is the sum of an invariant series which is even in $v$ and an invariant series which is odd in $v$. This follows immediately from the fact that if a series is invariant under $T$ it is also invariant under $T^{-1}$.

Furthermore, if the transformation $\mathfrak{T}$ is written in the variables $y$ and $x$ by means of the equations (16) it is readily seen that every series that is formally invariant is the sum of an invariant series which is even in $x$ and an invariant series which is odd in $x$.

Now let us consider the equations (24) for the transformation $\mathfrak{T}$. These equations may be written in the form

$$
\begin{aligned}
F(u,v) &= a_0(u) + a_1(u)v + a_2(u)v^2 + a_3(u)v^3 + \cdots \\
&= a_0(u_1) - a_1(u_1)v_1 + a_2(u_1)v_1^2 - a_3(u_1)v_1^3 + \cdots, \\
G(u,v) &= b_0(u) + b_1(u)v + b_2(u)v^2 + b_3(u)v^3 + \cdots \\
&= -b_0(u_1) + b_1(u_1)v_1 - b_2(u_1)v_1^2 + b_3(u_1)v_1^3 - \cdots .
\end{aligned} \tag{30}
$$

From the reasoning given in the preceding paragraph it is evident that if there is a formal series which is invariant under $T$ there is one, $K(y, x)$,

in the variables $y$ and $x$ which is invariant under $T$ and is even in $x$ and consequently in $G(u, v)$ when the variables $y$ and $x$ are replaced by the functions $F(u, v)$ and $G(u, v)$ of (16). Hence from (24),

$$K[F(u_1,v_1), G(u_1,v_1)] = K[F(u,v), G(u,v)] = K[F(u_1, -v_1), -G(u_1, -v_1)].$$

But $K$ is an even function of $G$ by hypothesis. Hence

$$K[F(u_1,v_1), G(u_1,v_1)] = K[F(u_1, -v_1), G(u_1, -v_1)],$$

whence $K$ is even in $v_1$, and hence, when it is expressed in terms of $u$ and $v$, even in $v$. In an analogous manner it can easily be proved that any invariant series which is odd in $x$ in the $yx$-system of coördinates is odd in $v$ when written in the $uv$-system of coördinates. Since the square of an invariant series which is odd in $v$ is an invariant series which is even in $v$ we see that there exists a series which is formally invariant under $T$ if and only if there exists such a series which is even in $v$.

We shall now examine transformations $\mathfrak{T}$ to find out whether formally invariant series can exist, and we shall limit the discussion at first to the cases in which the transformations are of the type V, or VI with $\alpha \neq 0$. Since we have shown that there exists a formally invariant series for such a transformation if and only if there exists one which is even in $x$, and hence, as has been shown, even in $v$, we shall try to show that, for the cases under consideration, there exists a formally invariant series which is even in $x$ and, when expressed in terms of $u$ and $v$, is even in $v$.

It is readily provable that if there exists an invariant series it can not have any terms of the first degree in $u$ or in $v$. Hence we shall start with second degree terms. Let us write the equations (30) in the form

$$\begin{aligned} F(u,v) &= a_{10}u + a_{01}v + a_{20}u^2 + a_{11}uv + a_{02}v^2 + \cdots \\ &= a_{10}u_1 - a_{01}v_1 + a_{20}u_1^2 - a_{11}u_1v_1 + a_{02}v_1^2 + \cdots, \\ G(u,v) &= b_{10}u + b_{01}v + b_{20}u^2 + b_{11}uv + b_{02}v^2 + \cdots \\ &= -b_{10}u_1 + b_{01}v_1 - b_{20}u_1^2 + b_{11}u_1v_1 - b_{02}v_1^2 + \cdots. \end{aligned} \tag{31}$$

Now from the way in which the series $F(u, v)$ and $G(u, v)$ are transformed by $T$ we see that our invariant series which is to be even in $x$ and consequently in $G$ must be such that when $F$ and $G$ are expanded in terms of $u$ and $v$ no term of odd degree in $v$ appears. In so far as the second degree terms are concerned, except for a constant factor, there is only one combination of $F$ and $G$ which is even in $v$ and invariant under $T$ up to terms of degree three, and this is

$$b_{10}b_{01}F^2 - a_{01}a_{10}G^2. \tag{32}$$

This expression transforms under $T$ in the same manner as do $F$ and $G^2$, and has no term of the second degree in $u$ and $v$ which is odd in $v$. The method which we shall use to build up our formally invariant series consists in adding to the function (32) a function of the third degree in $F$ and $G$ which is even in $G$ and which is such that the resulting sum contains no term of the third degree in $u$ and $v$ which is odd in $v$; and then adding to this sum a function of the fourth degree in $F$ and $G$ which is even in $G$ and which is such that this last sum will contain no term of the fourth degree in $u$ and $v$ which is odd in $v$; this process being continued so long as any terms of odd degree in $v$ remain in the sum. Now the series (32) is invariant up to its terms of the third order; the series obtained from (32), after the first addition has been made to it, will be invariant up to terms of the fourth order; and so on. Thus, if we can prove that at no stage does the above outlined process of adding functions break down, we shall have established the fact that there exists an invariant series for transformations $\mathfrak{T}$ of the types V and VI with $\alpha \neq 0$.

Let us now investigate what we need to prove in order to show that this process does not break down at any stage. Suppose we wish to eliminate terms of odd order in $v$ in the group of terms of the $r$th degree in $u$ and $v$. To do this we may employ a sum of the form

$$\alpha_0 G^r + \beta_0 G^{r-2}F^2 + \cdots + \eta_0 F^r \tag{33}$$

if $r$ is even, and of the form

$$\alpha_0 G^{r-1}F + \beta_0 G^{r-3}F^3 + \cdots + \eta_0 F^r \tag{34}$$

if $r$ is odd.

Let us first consider the case when $r$ is odd. When a sum of the form (34) is employed, we see that to determine the $(r+1)/2$ multipliers $\alpha_0, \beta_0, \cdots, \eta_0$, we shall have $(r+1)/2$ linear algebraic equations. Hence, if we are to show that in every case we shall be able to solve these equations we shall have to show that the determinant of the coefficients of $\alpha_0, \cdots, \eta_0$ in these equations is different from zero. Proving this is equivalent to proving that if the function (34) is even in $v$ up to terms of the $(r+1)$th order, $\alpha_0 = \beta_0 = \cdots = \eta_0 = 0$. Evidently, in this proof only the linear terms of $F$ and $G$ will be involved. Hence let us define

$$\begin{aligned} F_0(u,v) &= a_{10}u + a_{01}v, \\ G_0(u,v) &= b_{10}u + b_{01}v, \end{aligned} \tag{35}$$

where $F_0$ and $G_0$ are thus the linear terms of $F$ and $G$, respectively. Hence $a_{10}b_{01} - a_{01}b_{10}$ may be assumed to be 1. Consequently the transformation $\mathfrak{T}_0$,

which is defined by

$$T_0: \quad \begin{aligned} u_1 &= \alpha u + \beta v, \quad \alpha = a_{10}b_{01} + a_{01}b_{10}, \quad \beta = 2a_{01}b_{01}, \\ v_1 &= \gamma u + \alpha v, \quad \gamma = 2a_{10}b_{10}, \end{aligned} \tag{36}$$

and hence is the transformation obtained by using only the first degree terms in the equations $T$ of (26), is the transformation defined by

$$\begin{aligned} F_0(u_1, -v_1) &= F_0(u,v), \\ G_0(u_1, -v_1) &= -G_0(u,v). \end{aligned} \tag{37}$$

Since we are considering only transformations $\mathfrak{T}$ of the types V and VI in which $\alpha\beta\gamma \neq 0$, by means of a linear change of variables,

$$w = \beta^{1/2}v - \gamma^{1/2}u, \quad z = \beta^{1/2}v + \gamma^{1/2}u,$$

the equations of (36) may be written as

$$w_1 = \rho w, \quad z_1 = \frac{1}{\rho}z, \tag{38}$$

where now imaginary quantities may occur. For the argument to follow, it is necessary that $\theta$, where $\rho = e^{i\theta}$, be not a rational multiple of $2\pi$.

Any series which is invariant under the transformation (38) must be of the form

$$\sum_r c_r w^r z^r.$$

Hence, any series which is invariant under $T_0$ must be of the form

$$\sum c_r(\beta v^2 - \gamma u^2)^r,$$

which is the same as

$$\sum c_r 2^r (b_{10}b_{01}F_0^2 - a_{10}a_{01}G_0^2)^r. \tag{39}$$

This shows first of all that any series that is invariant under $T_0$ has only terms of even order in $u$ and $v$; and secondly, it shows that the terms of any degree form an integral power of $(b_{10}b_{01}F_0^2 - a_{10}a_{01}G_0^2)$. Thus any linear combination of terms of even degree in $F_0$ and $G_0$ such as

$$\alpha_0 G_0^{r-1}F_0 + \beta_0 G_0^{r-3}F_0^3 + \cdots + \eta_0 F_0^r, \tag{40}$$

$r$ being an odd integer, can not be invariant under $T_0$ unless $\alpha_0 = \beta_0 = \cdots = \eta_0 = 0$, which, on account of the definition of $T_0$ by the equations (36), means that if the function (40) is even in $v$, $\alpha_0 = \cdots = \eta_0 = 0$, since if it is even in $v$ it is invariant under $T_0$. Hence the process described above relative to adding functions of $F$ and $G$ to the function (32) to form an in-

variant series will never fail when we wish to eliminate terms of odd degree in $u$ and $v$ which are odd in $v$.

Now consider the case when the terms to be eliminated are of even degree in $u$ and $v$ and a series of the form (33) is employed. We see that to determine the $\alpha_0, \cdots, \eta_0$, we shall have to solve a system of $r/2$ linear algebraic equations in the $(r+2)/2$ unknowns $\alpha_0, \cdots, \eta_0$. If we choose $\alpha_0=0$ we shall have $r/2$ equations to solve for $r/2$ unknowns and it is evident from the reasoning for the case when $r$ is odd that the determinant of the coefficients of the unknowns will be always different from zero and that the equations can always be solved. This completes the proof that the process of building up a series which is formally invariant under $T$ when $T$ is of the type V or VI with $\alpha \neq 0$, and the condition following (38) satisfied, never fails.

Thus we may assert that when the transformation $\mathfrak{T}$ is of the type V or of the type VI with $\alpha \neq 0$ and the condition following (38) satisfied, there exists a series, $H^*(u, v)$, whose initial terms are $\beta v^2 - \gamma u^2$, which is formally invariant under $T$.

Let us now examine transformations $\mathfrak{T}$ of the type I(B) for invariant series, and let us use the same general plan that we used when $\mathfrak{T}$ was of the type V or VI. In this case we may choose $F$ and $G$ so that

$$F(u,v) = u + \tfrac{1}{2}v + \cdots,$$
$$G(u,v) = v + \cdots,$$

as has already been shown. Evidently the only quadratic function which is invariant under $T$ up to terms of the third order is a multiple of $G^2$.

Define $F_0 = u + \frac{1}{2}v$, $G_0 = v$, and the corresponding transformation $\mathfrak{T}_0$ has as its equations

$$T_0: \quad u_1 = u + v, \qquad v_1 = v,$$

and the only series which is formally invariant under $T_0$ is evidently a series in $v$ alone, and hence a series in $G_0$ alone. Thus, an argument similar to that given when $T$ was of the type V or VI shows that any series of the form

$$\alpha_0 G_0^{r-1} F_0 + \beta_0 G_0^{r-3} F^3 + \cdots + \eta_0 F_0^r, \tag{40}$$

where $r$ is an odd positive integer, is an even function of $v$ only if $\alpha_0 = \cdots = \eta_0 = 0$; and any series of the form

$$\beta_0 G_0^{r-2} F_0^2 + \cdots + \eta_0 F_0^r, \tag{41}$$

where $r$ is an even positive integer, is an even function of $v$ only if $\beta_0 = \cdots = \eta_0 = 0$. Hence, from reasoning similar to that given when $\mathfrak{T}$ was

of the type V or VI, it follows that every transformation of the type I(B) possesses a formally invariant series whose initial term is $v^2$.

Let us now consider transformations of the type II(B). The only series invariant under the corresponding $T_0$ are series in $-u$, i.e., $G_0$. Thus it is evident, from the type of reasoning given above, that for any transformation of this type there exists an invariant series whose first and second degree terms are all zero except the one involving $u^2$, which is not zero.

Now consider the possibility of invariant series for the cases where $\mathfrak{T}$ can be represented by equations of the form I(A), II(A), III or IV. Since series invariant under the corresponding equations $T_0$ may be series in $F_0$ only, we see that our method of proving the existence of invariant series fails since a series of the form (40) or (41) may be an even function of $v$ and still not have $\alpha_0 = \cdots = \eta_0 = 0$. However, on making use of the fact that our method does fail, we can actually set up transformations of these types which possess no formally invariant series. This is due to the fact that, for these cases, the determinant of the coefficients in the equations which determine the $\alpha_0, \cdots, \eta_0$ of (33) and (34) may now be zero, and by a proper choice of the coefficients of the $F$ and the $G$ these equations will be inconsistent. Hence we can only say that transformations of the types I(A), II(A), III and IV may or may not have formal invariant series.

We may say by way of summary that *for equations $T$ of the types* I(B), II(B), V *and* VI ($\alpha \neq 0$, *and $\theta$ satisfying the condition mentioned after* (38)), *there always exist formal invariant series; but for equations of the other types such series need not exist.*

## V. Totality of invariant series

For certain types of transformations $\mathfrak{T}$ we have proved the existence of formal invariant series and have given rules for finding particular ones. We now wish to prove that *for any given transformation every formal invariant series is a formal series in powers of a particular one.*

Let us first consider those transformations of types V and VI for which we proved the existence of invariant series. As we have seen, there is no invariant series with any terms of degree less than two. Hence, in every invariant series, the terms of lowest degree are of at least the second degree. Now suppose that the degree of the lowest degree terms of a formal invariant series is $r \geqq 2$, and suppose that these terms do not form a multiple of some power of

$$b_{10}b_{01}F_0^2 - a_{01}a_{10}G_0^2$$

of (32). Since the invariancy of these terms is due entirely to their properties

as functions of $F_0$ and $G_0$ it follows from the discussion relative to the equations (37), (38) and (39) that we have a contradiction. Hence every formal series invariant for a transformation of the type V, or of the type VI in which $\alpha \neq 0$ and $\theta$ satisfies the condition mentioned after the equation (38), is a formal series in the one whose existence was proved in the last section. This is seen as follows. Let us denote the invariant series whose existence was proved in the last section by $F^*$, and denote any other invariant series by $F$. Let $F_m^*$ and $F_m$ be the terms of minimum degree in $F^*$ and in $F$, respectively. Then we have that $F_m = c(F_m^*)^r$ for some positive integer $r$ and some constant $c$. Then is $F - c(F^*)^r$ an invariant series, and its terms of minimum degree have a degree greater than $r$ and form a multiple of a power of $F_m^*$. This series may now take the place of $F$ in the argument just given and that argument may be repeated.

Now consider transformations of the type I(B). We have shown that there exists a formal invariant series whose only term of degree two or less is $v^2$. On the other hand, if there exists a formal invariant series which has terms odd in $v$, it follows from the argument of the last section that there exists a formal invariant series which contains $v$ as a factor. In this case the line $v=0$ is an invariant curve, which, on account of the fact that the transformation $\mathfrak{T}$ is of the form $\mathfrak{RS}$, is a line of invariant points.

Conversely, if $v=0$ is a line of invariant points, the $u$ and the $y$ axes coincide, which implies that the function $G(u, v)$ of (16) has the form

$$G(u,v) = v[1 + G_1(u,v)], \tag{42}$$

where $G_1$ has no constant term. We wish now to show that in this case every invariant series has $v$ as a factor. An invariant series contains $v$ as a factor if and only if it contains $G$ as a factor since $G$ contains $v$ as a factor, so we shall have proved what we wish to prove if we can show that every invariant series has $G$ as a factor.

We shall prove first of all that there exists an invariant series which is odd in $v$ by using a method similar to that used in proving the existence of invariant series which are even in $v$. The function $G(u, v)$ is invariant under $T$ up to terms of the second degree and every term contains $v$ as a factor. Now under $T$, $G(u, v)$ and $F(u, v)$ are transformed according to the equations (31) where $b_{i0}=0$, $i=1, 2, \cdots$. Furthermore, every series in $F$ and $G$ which is odd in $G$ transforms under $T$ as $G$ itself does and contains $v$ as a factor. Hence we wish to add to $G$ a multiple of $FG$ so that the resulting sum up to terms of the third degree is invariant under $T$, and odd in $v$; then add a multiple of $F^2G$ so that the result will be invariant under $T$ up to terms of the fourth degree, and will be odd in $v$; and so on. At no stage do we introduce

terms in $u$ alone. When we add terms of the $r$th degree they are of the form

$$\alpha_0 F^{r-1}G + \beta_0 F^{r-3}G^3 + \cdots + \eta_0 F^{r-r^*}G^{r^*} \tag{43}$$

where $r^*$ is $r-1$ if $r$ is even, and is $r-2$ if $r$ is odd. Hence there are in (43) $r/2$ coefficients to determine if $r$ is even, and $(r-1)/2$ if $r$ is odd. But these coefficients are to be determined so that there will be no terms of degree $r$ in the series which are of even degree in $v$. Hence there will be $r/2$ linear algebraic equations to determine these coefficients if $r$ is even, and $(r-1)/2$ if $r$ is odd. From the reasoning given in the preceding section, it follows that the determinant of the coefficients of the $\alpha_0, \cdots, \eta_0$ in these equations is always different from zero so that they can be solved. Hence the above process of building up an invariant series having a linear term in $v$ and having $v$ as a factor fails at no stage. Hence there exists such a series which we shall denote by $S^*$.

Now consider any other series which is invariant under $T$. There is only one term of minimum degree and that is a multiple of a power of $v$. This follows at once from the form of the linear terms in the equations representing the transformation, $T$. Hence this formal invariant series is a formal series in $S^*$, for if it were not, we could add to it a formal series in $S^*$ and obtain a formal invariant series which has a term of minimum degree containing $u$ only, which is impossible. Hence, *if $v=0$ is a curve of invariant points for a transformation of the type* I(B), *every formal invariant series contains $v$ as a factor as many times as the degree of its terms of lowest degree.*

Thus we have shown that for transformations of type I(B), if there exists an invariant series which is odd in $v$, there exists one which is odd in $v$ and has a linear term in $v$, and has the further property that every other invariant series is a formal power series in this one. On the other hand, if there exists no invariant series which is odd in $v$, we still have existing one which is even in $v$ and has as initial term $v^2$. In this case, also, it can be shown that every other invariant series is a power series in this one.

In a similar manner it can be shown that for every transformation of the type II(B) every invariant series is a formal series in the one whose initial term is $u^2$.

Now let us turn our attention to transformations of the types I(A), II(A), III, IV and VI (cases not included in the argument at the first of the section). As has been noted, a transformation may be of one of these types and not have a formal invariant series. If a transformation is of one of these types and possesses an invariant series, it possesses one whose terms of minimum degree have a degree less than or equal to that of every other

invariant series. Then it can be proved by means of methods other than those used above that every formal invariant series is a formal series in one such as has just been described.†

## VI. Invariant curves

Let us fix our attention on any particular transformation $\mathfrak{T}=\mathfrak{R}\mathfrak{S}$, and consider any curve $C$ which passes through the origin. It is reflected by $\mathfrak{S}$ in the $y$ axis into a curve $C_s$ and $C_s$ is reflected by $\mathfrak{R}$ in the $u$ axis into a curve $C_{rs}=C_t$. If the curve $C$ is invariant under $\mathfrak{T}$, it is the same as $C_t$. But $C_t$ and $C_s$ are, in the $uv$ system of coördinates, the reflection images of one another in the $u$ axis. Hence, if $C=C_t$, we see that $C$ and $C_s$ are the reflection images of one another in the $u$ axis in the $uv$ system of coördinates as well as the reflection images of one another in the $yx$ system of coördinates. Hence, if $C$ is invariant under $\mathfrak{T}$, $C_s$ is also, and thus we see that invariant curves occur in pairs, each curve of every pair being the reflection of the other curve of the same pair in the $y$ axis under $\mathfrak{S}$ and in the $u$ axis under $\mathfrak{R}$. We shall call such a pair of invariant curves a "pair of conjugate curves." It may happen, of course, that a pair of conjugate invariant curves consists of only one curve counted twice.

Let us now discuss the possibility of there being curves of invariant points through the origin. If such curves exist, it follows from the equations (24) for the transformation $\mathfrak{T}$ that the coördinates of their points must satisfy the equations

$$F(u, -v) = F(u,v), \quad G(u, -v) = -G(u,v),$$

which implies that such curves are analytic. The above equations also show that either the origin is an isolated invariant point or there exists a curve of invariant points passing through it.

It is of interest to note that if the points on an invariant curve through the origin are not invariant, they either approach the origin under the transformation $\mathfrak{T}$ or recede from it. If the points on one invariant curve approach the origin under the transformation, the points on its conjugate recede from the origin. But a proof of this latter fact will be omitted since it is not used in any of the discussion to follow.

We now wish to discuss the relationships existing among formal invariant series, formal invariant curves through the origin, and actual invariant curves through the origin. In case there exists a formal invariant series,‡

† B, p. 25. Reference is to the first part of the theorem only. To prove our statements are correct for transformations of the types II(A) and IV, note that every series formally invariant under $T$ is also invariant under $T^2$.

‡ B, pp. 18–33.

$F(u, v)$, then every real formal curve which makes a factor of this series vanish is a formal invariant curve, and conversely, every formal invariant curve which is not made up of invariant points makes one of the factors of $F$ vanish. That a formal invariant curve consisting of invariant points does not necessarily make a factor of $F$ vanish will be shown to be the case by means of an example in the next section. It will be shown in this same section that there may exist formal invariant curves and yet no invariant series. This illustrates some of the radical differences between transformations of the type discussed in this paper and those of the type discussed by Birkhoff, for when a transformation is represented by equations whose linear terms are those of the type I(A), it possesses an invariant series if it is conservative, and every formal invariant curve, whether made up of invariant points or not, makes one of the factors of this series vanish.

It has been shown,† in an ingenious manner, that to every real formal invariant curve of a conservative transformation of any one of a number of types there corresponds one and only one actual invariant curve, which, moreover, is hypercontinuous at the origin which is the invariant point under consideration, and has the formal invariant curve as its asymptotic representation.

Furthermore, in any sector of a neighborhood of the origin, sufficiently small, which does not contain any invariant curve which has a formal invariant curve as its asymptotic representation at the origin, and which is bounded by two such invariant curves which are not curves of invariant points, no point remains on indefinite iteration of the transformation or of its inverse.

We wish now to outline proofs of analogous statements concerning general transformations whose representing equations are of the form

$$u_1 = u + \cdots, \qquad v_1 = v + \cdots, \tag{44}$$

where the dots indicate terms of degree greater than one. It will be assumed that the equations (44) possess formal invariant series. The formal invariant series used in the discussion is one whose leading terms are of degree not greater than that of the leading terms of every other formal invariant series. It will be recalled that every formal invariant series is a power series in such a one. A comparatively simple case will be treated first in order to illustrate the method, which is a modification of that used by Birkhoff to prove the theorems just mentioned. We shall not consider the case where the formal invariant curves are curves of invariant points since such curves are analytic.

† B, Chapter II.

We shall suppose that the real formal invariant curve under consideration is regular at the origin, i.e., can be represented in the form $u=$a power series in $v$, or in the form $v=$a power series in $u$. Let us suppose that our curve is given in the latter form and that its equation is

$$v = p(u) = a_1u + a_2u^2 + \cdots . \tag{45}$$

If we perform the formal change of coördinates

$$u^* = u, \qquad v^* = v - p(u), \tag{46}$$

the formal curve (45) is taken as the $u^*$ axis and the equations of the transformation are reduced to the form

$$\begin{aligned} u_1^* &= u^* + \cdots , \\ v_1^* &= v^*(1 + \cdots ), \end{aligned} \tag{47}$$

and the transform of the formal invariant series used has $v^*$ as a factor, i.e.,

$$F(u,v) \sim v^*( \cdots ). \tag{48}$$

In the series on the right hand side of the first equation of (47) let $au^{*p}$ be the first term, other than $u^*$, which does not contain $v^*$ as a factor. There is such a term, for otherwise $v^*=0$ is a formal curve of invariant points which is contrary to supposition. For definiteness in argument we shall suppose that $a>0$.

There are two distinct cases to consider, the first being where the formal invariant series of (48) in $u^*$ and $v^*$ contains a term containing $u^*$, and the other the case where the formal invariant series contains no such term. We shall now consider the first case.

Let the equations† which represent the $k$th iterate of this transformation be

$$u_k^* = \sum_{m+n=1}^{\infty} \phi_{mn}^{(k)} u^{*m} v^{*n},$$

$$v_k^* = \sum_{m+n=1}^{\infty} \psi_{mn}^{(k)} u^{*m} v^{*n},$$

and define

$$\delta u^* = \left.\frac{\partial u_k^*}{\partial k}\right|_{k=0}, \quad \delta v^* = \left.\frac{\partial v_k^*}{\partial k}\right|_{k=0}, \tag{49}$$

whence $\delta u^*$ and $\delta v^*$ are formal series in $u^*$ and $v^*$. Now the formal series $F$ satisfies the equation

† B, p. 10.

$$\frac{\partial F}{\partial u^*}\delta u^* + \frac{\partial F}{\partial v^*}\delta v^* = 0,$$

and the first term in $\delta u^*$ which is independent of $v^*$ is $au^{*p}$.

Let $F$ contain $v^{*l}$ and no higher power of $v^*$ as a factor. Then $\partial F/\partial u^*$ can be factored into the form

$$\frac{\partial F}{\partial u^*} = v^{*l}(u^{*k-1} + A_1 u^{*k-2} + \cdots + A_{k-1})(c_1 + \cdots),$$

where $c_1$ is a constant different from zero and the $A_i$ are zero for $v^*=0$. Using analogous notations we have

$$\frac{\partial F}{\partial v^*} = v^{*l-1}(u^{*k} + \cdots)(c_2 + \cdots),$$

$$\delta u^* = (u^{*p} + \cdots)(c_3 + \cdots),$$

$$\delta v^* = v^{*s}(u^{*q} + \cdots)(c_4 + \cdots).$$

From the differential equation displayed after the equations (49) we have immediately that there is a term in $\delta v^*$ of the form $cv^*u^{*p-1}$ where $ac<0$, and there is no term in $\delta v^*$ of lower degree in $u^*$ which does not contain at least $v^{*2}$. But this implies that the series for $v_1^*$ contains the term $cv^*u^{*p-1}$ and there is no other term of this series of the form $dv^*u^{*m}$ where $m<k-1$.

The formal invariant series contains a term $bv^{*l}u^{*k}$. The change of coördinates (46) is not analytic, but if we replace the series $p(u)$ by $p_n(u)$ where $n$ is arbitrarily large and larger than $l+k+2p$, the $p_n(u)$ being the sum of the first $n$ terms of $p(u)$, the new change of coördinates is analytic and the new equations of the transformation agree with the formal equations (47) out to terms of the $n$th degree. Assuming such a transformation has been made, dropping the asterisks for simplicity in notation, we may write the equations of our transformation, and the formal invariant series, as

$$u_1 = [u + \cdots + au^p + \cdots] + \cdots, \tag{50}$$

$$v_1 = v[1 + \cdots + cu^{p-1} + \cdots] + \cdots,$$

$$F = v^l[\cdots + bu^k + \cdots] + \cdots, \tag{51}$$

where the expression inside the brackets in the first equation of (50) is a polynomial of degree $n$ at most, those inside the other brackets in (50) and (51) are polynomials of degree $n-1$ at most, and all the terms outside the brackets are of degree at least $n+1$.

We have now the hypotheses necessary for the application of the results of Birkhoff. We shall quote them without going into the details of the proofs.

There exists a positive integer $n_0$ such that for $n>n_0$ there exists a constant $d$ and a closed set $\Sigma$ of curves defined for $u$ sufficiently small and positive, analytic for $u>0$, which is invariant under $\mathfrak{T}$ and lies within the region defined by

$$0 < u < d, \qquad |v| < Eu^{n-n_0},$$

where $E$ is a positive constant depending on $n$ and $d$. Define $m=n-n_0$, which is greater than $p$ due to the choice of $n$. These curves have contact of order at least $m$ at the origin with the $u$ axis as chosen in (50). Now consider the region which is bounded by these curves of $\Sigma$ and which contains all of them. The upper and lower boundaries of this region are invariant curves for the transformation under consideration and have an order of contact of at least $m$ at the origin with the $u$ axis. We wish now to show that these are one and the same curve, and hence, that $\Sigma$ consists of only one curve.

Since the distance of any point $(u, v)$ on either of these boundary curves from the $u$ axis is of the order $|u|^m$, by making a transformation which takes one of them and its analogue for $u<0$ into the new $u$ axis, we may write the transform of (50) as

$$(52) \qquad \begin{aligned} u_1 &= [u + au^p] + \phi_1(u,v), \\ v_1 &= v\{[1 + cu^{p-1}] + \phi_2(u,v)\}, \end{aligned}$$

where $\phi_1(u, v)=O(u^q)$ and $\phi_2(u, v)=O(u^{q-1})$, for some positive integer $q>p$ if $|v|<c_0|u^p|$, where $c_0$ is an arbitrary constant. From these facts and the equations (52) it follows that for $u$ small enough in absolute value, and $|v|<c_0|u^p|$, the ratio $(v_1-v)/(u_1-u)$ is negative if $u$ and $v$ are positive, positive if $u$ is positive and $v$ is negative, etc., since $ac<0$. This shows that there can not be another invariant curve passing through the origin in this restricted region, besides that which we have chosen as our new $u$ axis. Hence $\Sigma$ consists of exactly one curve and is analytic except possibly at the origin.

But this $u$ axis has contact of at least the $m$th order with the original one used in (50). Hence any curve which has contact of order greater than $m$ with the $u$ axis of (50) has contact of order $m$ at least with the $u$ axis as chosen in the equations (52), and, since $m>p$, must lie in the above mentioned region for $u$ small enough in absolute value, and for some value of $c$. Hence the invariant curve that we have shown is the only curve of $\Sigma$, is analytic for $(u, v)\neq(0, 0)$, and has contact of at least order $m$ with the $u$ axis of the equations (50), is that which will be obtained if we choose higher and higher values for $n$ mentioned in connection with the equations (50). This completes the proof that *if $F$ of* (48) *contains a term having $u^*$ as a factor, to each real formal invariant curve which is regular at the origin there corresponds exactly*

*one actual invariant curve which has this formal invariant curve as its asymptotic representation at the origin.*

Let us now consider what can be said about the series for $v_1$ in case the formal invariant series $F$ in (48) contains no term in $u^*$. In this case there is only one real formal invariant curve which is not a curve of invariant points, and when that has been transformed formally into the $u^*$ axis the expression for $v_1^*$ is a power series in $v^*$ alone. But if the series $F$ has no term involving $u^*$ its term of minimum degree is a multiple of a power of $v^*$, and, since the degree of this term is not greater than that of the corresponding term in every other formal invariant series, it is one. But if there is a formal invariant series whose leading term is $v^*$ which contains no term involving $u^*$, there is one whose only term is $v^*$. Hence there is a formal invariant series for the transformation represented by the equations (44) of the form $v-p(u)$ where $p(u)$ is a power series in $u$.

Let us now recall that we intend to apply the discussion of this section to transformations of the type $\mathfrak{T}=\mathfrak{RS}$ where $\mathfrak{R}$ and $\mathfrak{S}$ are each of the reflection type, and we shall assume that the coördinate system used is a normal one for the transformation $\mathfrak{R}$. Under these assumptions, if there exists a formal invariant series of the form $v-p(u)$ there exists one of the form $v+p(u)$ which is implied by the discussion of Section IV, and this implies that $v$ is an invariant function which, in turn, implies that the second equation of (44) has the form $v_1=v$. In this case $v=0$ is an invariant curve. Hence, *when a transformation $\mathfrak{T}=\mathfrak{RS}$ is represented by equations of the form* (44) *where the uv system of coördinates is normal for $\mathfrak{R}$, and if there is a formal invariant series for this transformation which can be formally reduced to a formal series in $v^*$ alone by means of a change of variables of the form* (46), $v=0$ *is an invariant curve and, on account of the fundamental properties of $\mathfrak{T}$ as the product of $\mathfrak{R}$ and $\mathfrak{S}$, it follows that $v=0$ is a curve of invariant points.*

Now let us consider the transformations whose representative equations have the form (44) but whose formal invariant curves have "cusps" at the origin. By use of transformations of the type

$$u = u^*v^*, \qquad v = v^*, \tag{53}$$

such formal invariant curves are taken into ones of the first type considered in the above argument, i.e., the reduced equations representing the transformation have not a formal invariant series of the form $v^*-p(u^*)$. This is due to the fact that on account of the nature of the change of variables (53) there is one formal invariant series $F^*$ which contains either $u^*$ or $v^*$ as a factor, and to the fact that, according to Section V, every other formal invariant series is either a root or a formal power series in $F^*$.

Hence, *if there exists a real, formal, invariant series, to every real formal invariant "cusp," there corresponds one and but one real invariant curve which has this formal invariant curve as its asymptotic representation.*

Finally, concerning the stability of the invariant point $(u, v) = (0, 0)$ under a transformation represented by equations of the form (44), it may be said that *in case a formal invariant series exists, in any sector of a sufficiently small neighborhood of the origin, which does not contain any invariant curve which has a formal invariant curve as its asymptotic representation at the origin, and which is bounded by two such invariant curves which are not curves of invariant points, no point remains on indefinite iteration of the transformation or of its inverse.* The proof of this statement is omitted here since it is almost exactly that given in the discussion of conservative transformations which has been mentioned before.

## VII. Transformations of type I(A)

One of the chief differences between conservative transformations and those of the type studied in this paper is that for conservative transformations† of type II′′, which correspond to transformations of the type I(A), there always exist real formal invariant series, while such series need not exist for transformations of the type I(A). Due partially to the fact that formal invariant series need not exist, a wide variety of situations occur which are interesting in comparison with the facts known concerning conservative transformations of the type II′′. These situations will now be studied in some detail, while the analogous ones which arise for the other types of transformations for which there may not be any formal invariant series will hardly be mentioned since the discussion for type I(A) will be typical of that which could be given for them too.

We shall confine our attention to a neighborhood of the origin of coördinates which is supposed to be an invariant point. Questions which have to be answered are those relative to the existence of invariant curves through the origin, existence of invariant integrals, existence of invariant series, and stability.

Let us first note that in Section III we have already shown that all directions through the origin are invariant under a transformation of the type I(A). We shall now show that *for transformations of type* I(A), (i) *the existence of a formal invariant integral implies the existence of a formal invariant series*; (ii) *the existence of a formal invariant series does not imply the existence of a formal invariant integral*; (iii) *the existence of invariant curves does not imply the existence of formal invariant series*; *and* (iv) *the existence*

† See last of Section II.

*of formal invariant series which possess real factors*† *implies the existence of real invariant curves through the origin.* The proof of (i) is already in the literature.‡

We shall prove (ii) by means of an example. Consider the transformation defined by the following equations:

$$R: \quad u_1 = u, \quad v_1 = -v; \qquad S: \quad u_1 = u\frac{1-v}{1+v}, \quad v_1 = -v; \tag{54}$$

$$T = RS: \quad u_1 = u\frac{1-v}{1+v}, \quad v_1 = v.$$

Each of the transformations defined by the equations $R$ and $S$ is of the reflection type and it is evident that any series in $v$ is formally invariant under $T$. We need only show that there exists no quasi-invariant series. Evidently

$$\frac{\partial(u_1, v_1)}{\partial(u, v)} = \frac{1-v}{1+v}.$$

Now if $Q(u, v)$ is a quasi-invariant series we have

$$Q(u, v) = Q(u_1, v_1)\frac{\partial(u_1, v_1)}{\partial(u, v)},$$

and we must remember that $Q(0, 0)$ is different from zero. Let

$$Q(u, v) = q_{00} + q_{10}u + q_{01}v + \cdots.$$

Then

$$Q(u_1, v_1)\frac{\partial(u_1, v_1)}{\partial(u, v)} = (q_{00} + q_{10}u_1 + q_{01}v_1 + \cdots)\left(\frac{1-v}{1+v}\right)$$
$$= q_{00} + q_{10}u + (q_{01} - 2q_{00})v + \cdots.$$

Since $q_{00} \neq 0$, this equation shows that no formal quasi-invariant series $Q$ can exist for the transformation represented by (54).

We shall now show by an example that a transformation of the type I(A) may possess an invariant curve through the origin and yet possess no formal invariant series. Consider the transformation for which the equations (16) are

$$y = u + v^3, \quad x = v + u^3. \tag{55}$$

† B, p. 18.

‡ B, p. 16. In this paper, the assumption is made that the transformation possess an actual invariant integral, but for the proof of (i) no use is made of the fact that the quasi-invariant series $Q$ is convergent.

Evidently

$$x^2 - y^2 = (v^2 - u^2)(1 - 2uv - v^4 - v^2u^2 - u^4).$$

Hence $x^2-y^2=0$ is equivalent to $v^2-u^2=0$. But the curves $x^2-y^2=0$ are the images of one another under $\mathfrak{S}$ and the same curves $v^2-u^2=0$ are the images of one another under $\mathfrak{R}$. Hence the curves $x^2-y^2=0$, which are the same as the curves $v^2-u^2=0$, are a pair of conjugate invariant curves through the origin for the transformation $\mathfrak{T}=\mathfrak{RS}$.

Now turn to the equations (55) and investigate the possibility of the existence of a formal invariant series. We have seen that if such a series exists at all, there exists one that is even in $v+u^3$ and which, when expanded in terms of $u$ and $v$, is even in $v$. Furthermore, if there exists such a series, there exists one whose lowest degree terms are of even degree in $(v+u^3)$ and $(u+v^3)$. Let the lowest degree terms of such a series be

$$a_r(v + u^3)^r + a_{r-2}(v + u^3)^{r-2}(u + v^3)^2 + \cdots + a_0(u + v^3)^r, \tag{56}$$

where $r$ is an even positive integer. In this series (56) there exists no term of degree $r+1$ in $u$ and $v$. Furthermore, in our invariant series, when the terms of degree $r+1$ in $(v+u^3)$ and $(u+v^3)$ are expanded, there will be no terms of degree $r+2$. Hence the terms of (56) which are of degree $r+2$ in $u$ and $v$ but odd in $v$ have to be counterbalanced as in Section IV, by the addition of a sum of the form

$$b_0(v + u^3)^{r+2} + b_2(v + u^3)^r(u + v^3)^2 + \cdots + b_{r+2}(u + v^3)^{r+2}. \tag{57}$$

But there exist no terms in this series of degree $r+2$ which are of odd degree in $v$, and since the terms of (56) of degree $r+2$ which are of odd degree in $v$ are

$$\begin{array}{lllll} & & ra_rv^{r-1}u^3 & & \\ + 2a_{r-2}v^{r-1}u & & & + (r-2)a_{r-2}v^{r-3}u^5 & \\ & + 4a_{r-4}v^{r-1}u^3 & & & + (r-4)a_{r-4}v^{r-5}u^7 \\ & & + 6a_{r-6}v^{r-3}u^5 & & + \cdots \end{array}$$

. . . . . .

it is evident that $a_{r-2}=a_{r-6}=a_{r-10}=\cdots=0$ and $ra_r+4a_{r-4}=0$, $(r-4)a_{r-4}+8a_{r-8}=0, \cdots$. But since $a_{r-4\alpha}=0$ where $4\alpha$ is the largest multiple of 4 which is not greater than $r$, we see that each of this last set of $a_i$ is also zero, and hence there can exist no invariant series, although there exist analytic invariant curves through the origin.

In connection with transformations of the type I(A) it is of interest to know when there are curves of invariant points through the origin. Let the equations (16) be denoted by

$$y = u + f(u,v), \quad x = v + g(u,v),$$

where $f$ and $g$ are convergent series in $u$ and $v$ which have no terms of degree less than two. Then the transformation $\mathfrak{T}$ is represented by the equations

$$u_1 + f(u_1, -v_1) = u + f(u,v),$$
$$-v_1 + g(u_1, -v_1) = -v - g(u,v).$$

The points that are invariant under $\mathfrak{T}$ are given by

$$u + f(u, -v) = u + f(u,v), \qquad -v + g(u, -v) = -v - g(u,v),$$

i.e., $f(u,-v) = f(u, v)$, $g(u,-v) = -g(u, v)$. This means that $f$ is an even function of $v$ and $g$ is an odd function of $v$ for $(u, v)$ an invariant point. Hence, if $f$ and $g$ are such that for curves passing through the origin these equations are satisfied, these curves are curves of invariant points.

We have given examples of transformations under which certain curves passing through the origin are invariant. On the other hand, there exist transformations of the type I(A) under which no curve through the origin is invariant. An example of such a one is the transformation for which the equations (16) are

$$y = u + v^3, \quad x = v - u^3,$$

but we shall not pause to prove this fact.

In the study of conservative transformations the fact was discovered that, for each transformation of type II'', every invariant curve that passes through the origin and is hypercontinuous there has an asymptotic expansion at the origin which formally makes every formal invariant series zero, and formal invariant series always exist. We have already shown that for the analogous transformations of type I(A) there may exist analytic invariant curves through the origin and yet no formal invariant series; now we shall give an example of a transformation of type I(A) which has an invariant series and an analytic invariant curve through the origin along which the series is not formally zero. This invariant curve is necessarily a curve of invariant points.†

Consider again the transformation represented by

$$u_1 = u(1 - v)/(1 + v), \quad v_1 = v.$$

† B, pp. 31–32.

Any formal series in $v$ is a formal invariant series. Yet $u=0$ is a line of invariant points and $u$ is not a factor of any formal invariant series.

Finally, in connection with the questions relative to actual invariant curves corresponding to real factors of a formal invariant series, and questions of stability, we shall merely draw attention to the fact that the discussion of Section VI is applicable in this one.

## VIII. Transformations of type I(B)

We have already shown that for such transformations there always exists a formal invariant series which is even in $v$, whose only term of degree less than three is $v^2$. Hence this formal invariant series can be factored in the form

$$(v^2 + f(u))(1 + \cdots)$$

where $f(u)$ is a formal series in $u$ whose term of lowest degree we shall denote by $au^p$ in case $f(u)$ is not identically zero. As has already been seen in Section IV, $f(u)$ is identically zero if and only if $v=0$ is a line of invariant points.

On the other hand, if $f(u)$ is not identically zero, $v^2+f(u)=0$ represents a real, formal, invariant curve with a "cusp" at the origin if $p$ is odd, and two real, formal, invariant curves tangent at the origin if $a<0$ and $p$ is even. If $a>0$ and $p$ is even, there is no real, formal, invariant curve through the origin.

A change of coördinates of the type $u=u^*$, $v=u^*v^*$ reduces the equations representing a transformation of the type I(B) to equations of the type (44), and hence the methods of Section VI show that, when there is a formal invariant "cusp," there is a unique actual invariant curve having this formal "cusp" as its asymptotic representation at the origin, and when there are two real formal invariant curves through the origin there are exactly two actual invariant curves tangent to the $u$ axis at the origin and having these formal curves as their asymptotic representations.

Furthermore, when invariant curves through the origin exist, the origin is an unstable hyperbolic point, and the argument at the end of Section VI shows that if a sufficiently small neighborhood of the origin be chosen, points not on the invariant curves in this neighborhood do not remain in it on iteration of the transformation or of its inverse.

With transformations of the type I(B) there are connected formal differential equations of the type (49) which were used in the discussion of equations of the type (44). But for such a transformation the highest common factor of $\delta u$ and $\delta v$ is $v$, and there is a curve of invariant points only if there exists an invariant series having $v^2$ as a factor. Hence,† transformations of the type I(B) are formally conservative.

† B, p. 19 et seq.

## IX. Transformations of types II(A) · · · IV

For all these transformations, except those of type II(B), there may or may not exist formal invariant series. For transformations of the type II(B), a discussion analogous to that given for transformations of type I(B) can be given. It may be noted, however, that in this case there can be no curve of invariant points through the origin. There may, however, exist a curve through the origin which is invariant under the transformation and which consists of points all of which are invariant under $\mathfrak{T}^2$.

As has already been mentioned, a discussion similar to that given for transformations of the type I(A) may be given here for each of the other transformations named in this section. This we shall not do, but shall content ourselves with merely pointing out a few of the main distinctions.

For transformations of the type II(A) the origin is an isolated invariant point, and the points on every invariant curve through the origin transform under $\mathfrak{T}$ into points, each of which is separated on the invariant curve from its image by the origin. This is a direct consequence of the form of the linear terms in the equations for $T$ in (25). The discussion of Section VI is evidently applicable to the second iterate of a transformation of the type II(A).

If a transformation is of the type III, the origin is an isolated invariant point unless there is a single curve of invariant points through it. This is shown as follows. Let the equations (16) be denoted by

$$y = u + f(u,v), \quad x = \tfrac{1}{2}u + v + g(u,v),$$

where all the terms of $f$ and $g$ are of degree higher than two. The invariant points are given by

$$f(u, -v) = f(u,v), \quad -u = g(u, -v) + g(u,v).$$

The form of these equations at once proves that the statement above is correct.

The origin, for transformations represented by equations of the types II(B), III and IV of (25), is an unstable invariant point, and may be either hyperbolic or elliptic.

## X. Transformations of type V

If a transformation is of type V, by means of a real linear change of variables, equations representing it may be chosen which have the form

$$(58) \quad u_1 = \rho u + \cdots, \quad v_1 = (1/\rho)\, v + \cdots, \quad \rho = \alpha \pm (\alpha^2 - 1)^{1/2} \neq 0,$$

where now the $uv$ system of coördinates is not normal for $\mathfrak{R}$. Any transformation represented by equations of this form in which the right hand members

are analytic possesses† two real analytic invariant curves through the origin whose equations may be written in the form $u=f(v)$, $v=g(u)$. When these curves are taken as axes, it can be readily shown that the points on one of these curves in a neighborhood of the origin move into the origin on iteration of the transformation, while the points on the other curve move away from the origin. In the case where $\rho<0$ the origin always lies between a point and its image when these points lie on one of the invariant curves through the origin.‡ Furthermore, a neighborhood of the origin can be taken so small that any point in it not on one of these invariant curves through the origin moves out of it on iteration of the transformation or of its inverse. Hence, transformations of the type V are hyperbolic and unstable.

We now wish to point out the interesting facts that *there is a formal equivalence existing between transformations of the type* V *and conservative transformations of the types* I′ *and* I″. We shall first prove that every transformation of the type I′ or of the type I″ is formally of the type V and hence is formally the product of two transformations each of the reflection type.§ Now the equations representing a conservative transformation of the type I′ or of the type I″, by a formal change of variables, may be given either the form¶

$$U_1 = \rho U e^{cU^iV^i}, \quad V_1 = \frac{1}{\rho} V e^{-cU^iV^i}, \quad c \neq 0, \tag{59}$$

or the form

$$U_1 = \rho U, \quad V_1 = \frac{1}{\rho} V. \tag{60}$$

But the equations (59) are the product $RS$ of the two systems

$$R: \begin{aligned} U_1 &= \rho V e^{cU^iV^i}, \\ V_1 &= \frac{1}{\rho} U e^{-cU^iV^i}; \end{aligned} \qquad S: \begin{aligned} U_1 &= V, \\ V_1 &= U, \end{aligned}$$

each of which is of the "reflection" type. Evidently the equations (60) are also the product of two systems each of the "reflection" type.

We shall now prove that, conversely, the equations representing a transformation of type V are, by a formal change of variables, equivalent to equations of the form (59) or (60), and that every transformation of type V is

† Poincaré, Œuvres, vol. I, p. 202.

‡ Cf. Section III.

§ This fact was kindly pointed out to the author by Professor Birkhoff.

¶ B, pp. 34, 55.

formally conservative. Let us assume that the equations representing the transformation of type V under consideration are in the form (58). From our previous discussion we know that there exists a formal invariant series whose only term of degree less than three is $cuv$ where $c$ is a constant different from zero. When $\rho>0$, we know that there exists† a formal quasi-invariant series and hence that our transformation is formally conservative and that its representative equations may be formally reduced to the form (59) or to the form (60). When $\rho<0$, the equations for $\mathfrak{T}^2$ are of the same form as (58) with $\rho$ replaced by $\rho^2$, and hence represent a formal conservative transformation. Hence they can be reduced to the form (59) or to the form (60) and hence‡ the transformed equations of $\mathfrak{T}$ have the same form.

## XI. Transformations of type VI

Transformations of type VI where $\rho$ satisfies the condition mentioned after the equations (38) are formally conservative, and transformations which are conservative and of the type II′ are formally the product of two transformations each of the reflection type. We shall not give proofs of these statements since they are wholly similar to those of the analogous statements made in the previous section.

On account of the form of the initial terms in the invariant series which we proved to exist for transformations of this type, it is evident that no formal invariant curves exist through the origin. In fact, the discussion given in Section III shows that there can exist no invariant curve of any kind through the origin since no slope through the origin is invariant. Hence the origin is an elliptic invariant point which raises many important and interesting questions. It will have been observed that the questions answered in the latter part of this paper have been, to a great extent, concerned with hyperbolic invariant points. The author hopes to study the elliptic cases in the near future.

† B, pp. 19–20.
‡ B, p. 55.

Cornell University,
Ithaca, N. Y.

# ON GALOIS FIELDS OF CERTAIN TYPES*

BY
LEONARD CARLITZ

## 1. Introduction

Several writers have considered the relations between the $\zeta$-functions of an algebraic field and some of its subfields. Thus Artin† has, in a particular case, considered the question of the divisibility of the $\zeta$-function of a field by that of a subfield. In another paper‡ he has shown how all possible $\zeta$-relations can be found. Explicit results of a general nature are however not arrived at. Herglotz§ has investigated fields formed by the composition of several quadratic fields, thus generalizing a well known result of Dirichlet's.¶ Pollaczek‖ has obtained results of a similar kind for Abelian fields with group of type (1, 1).

In all the cases cited, use is made of Hecke's** functional equation for the $\zeta$-function in an arbitrary field. Thus, for example, in Artin's first paper, a $\zeta$-relation is proved, except for a finite number of factors, by quite elementary methods; employing the functional equation, it is seen to hold in its entirety. Again, Herglotz and Pollaczek deduce discriminantal relationships by means of the functional equation.

In the following an explicit relation between $\zeta$-functions is deduced. The fields considered include as special cases those of Artin (first paper), Herglotz, and Pollaczek. No use is made of the Hecke functional equation; instead a method of an elementary nature is employed. Relations between discriminants also are easily proved by direct means. The first result of interest may be formulated thus:

Let $K$ be an (absolute) Galois field of degree $m$ and group $G_m$.†† We make

* Presented to the Society, April 18, 1930; received by the editors in March, 1930.

† *Ueber die Zetafunktionen gewisser algebraischer Zahlkörper*, Mathematische Annalen, vol. 89 (1923), p. 147.

‡ *Ueber eine neue Art von L-Reihen*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 3 (1924), p. 88.

§ *Ueber einen Dirichletschen Satz*, Mathematische Zeitschrift, vol. 12 (1922), p. 255.

¶ Werke, vol. 1, p. 533.

‖ *Ueber die Einheiten relativ-abelscher Zahlkörper*, Mathematische Zeitschrift, vol. 30 (1929), p. 520.

** *Ueber eine neue Anwendung der Zetafunktionen auf die Arithmetik der Zahlkörper*, Göttinger Nachrichten, 1917, p. 90.

†† Hilbert, *Die Theorie der algebraischen Zahlkörper*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 4 (1894–95), p. 248.

the following assumptions concerning $G_m$: $G_l$ is an invariant subgroup of index $a$, $m=la$. The complex $(G_m-G_l+l\cdot I)$, $I$ representing the identity, may be exhibited as the sum of $l$ non-overlapping conjugate groups. Further let $k_a$ be the Galois field corresponding* to $G_l$. To the $l$ groups of order $a$ mentioned above, let there correspond the $l$ conjugate fields of degree $l$, $k_l$, $k_l', \cdots, k_l^{(l-1)}$. With the above definitions and assumptions we shall prove (§3)

$$\zeta^a\zeta_K = \zeta_{k_a}\zeta_{k_l}{}^a. \tag{1}$$

If we next suppose that $K$ is a *relative* Galois field, the base field being some $F$, while all the remaining assumptions on $K$ are taken over for $K/F$, then (1) becomes

$$\zeta_F{}^a\zeta_K = \zeta_{k_a}\zeta_{k_l}{}^a. \tag{1a}$$

(It should be noted that all our $\zeta$-relations are proved only for a half-plane.)

The second general result concerns the discriminants of the fields defined. If by $d(k)$ we denote the discriminant of any field $k$, we find (§4)

$$d(K) = \pm\, d(k_a)d^a(k_l), \tag{2}$$

when $K$ is absolute Galoisian. In the more general case, we may suppose that each $d$ in (2) is a relative with respect to the field $F$. We then readily derive

$$d^a(F)d(K) = \pm\, d(k_a)d^a(k_l), \tag{2a}$$

where each $d$ is now an absolute discriminant.

In §5 are sketched the proof of formulas like (1) and (2) for the case of Abelian fields of type $(1, 1, \cdots, 1)$; in particular might be mentioned (18) and (19).

In §§6, 7, (1) and (2) and their analogs of §5 are applied to prove results that appear new. The results of the first section are contained in (30) and (31). A typical though particular result of §7 follows:

Let $k_4{}^1=k([a+b^{1/2}]^{1/2})$, $k_4{}^2=k([2a+2(a^2-b^2)^{1/2}]^{1/2})$, $k_2=k([(a^2-b)b]^{1/2})$, where neither $a^2-b$ nor $(a^2-b)b$ is a square. Then $K$, the field compounded of $k_4{}^1$, $k_4{}^2$, $k_2$ is readily seen to be Galoisian of degree eight,† and we shall prove

$$\zeta^2\zeta_K = \zeta_{k_2}\zeta_{k_4{}^1}\zeta_{k_4{}^2}$$

and

$$d(K) = \pm\, d(k_2)d(k_4{}^1)d(k_4{}^2).$$

* Hilbert, loc. cit., p. 250.

† Seidelmann, *Die Gesamtheit der kubischen und biquadratischen Gleichungen mit Affekt bei beliebigem Rationalitätsbereich*, Mathematische Annalen, vol. 78 (1918), p. 232.

By means of (1) and (2) a ratio of class numbers is transformed into a corresponding ratio of regulators. In §8 this ratio of regulators is considered and, in a particular case, an upper and a lower bound of a simple sort are determined.

I wish to take this opportunity to acknowledge my indebtedness to Professor H. H. Mitchell for his valuable suggestions and his very helpful criticism.

## 2. Properties of $G_m$

Let $G_a$ be one of the $l$ subgroups of $G_m$ contained in $(G_m - G_l + l \cdot I)$. Its operators will be denoted by $T_1, \cdots, T_a$. The operators of $G_l$ will be taken as $S_1, \cdots, S_l$. We may suppose $S_1 = T_1 = I$.

Lemma 1.

$$G_m = G_l \cdot T_1 + \cdots + G_l \cdot T_a = G_l \cdot G_a, \tag{3}$$

*and the factor group* $G_m/G_l = G_a$.

To prove this, we observe first that the number of elements in $G_l \cdot T_1 + \cdots + G_l \cdot T_a$ is $la = m$. Further, no two complexes $G_l T_\alpha$, $G_l T_\beta$ have an element in common. For suppose

$$S_i T_\alpha = S_j T_\beta \quad (S_i, S_j \text{ in } G_l);$$

then

$$S_j^{-1} S_i = T_\beta T_\alpha^{-1}.$$

But $G_l$ and $G_a$ have only $I$ in common; accordingly

$$S_i = S_j \quad \text{and} \quad T_\alpha = T_\beta,$$

which shows that all the elements in the sum are distinct.

Lemma 2.

$$l \equiv 1 \pmod{a}. \tag{4}$$

1. Let $S_i (2 \leqq i \leqq l)$ be any element of $(G_l - I)$. Then the set $T_\alpha S_i T_\alpha^{-1}$ $(\alpha = 1, \cdots, a)$ consists of $a$ distinct elements of $G$. For from $T_\alpha S_i T_\alpha^{-1} = T_\beta S_i T_\beta^{-1}$ we have $S_i T_\alpha^{-1} T_\beta S_i^{-1} = T_\alpha^{-1} T_\beta$. But this implies $T_\alpha = T_\beta$.

2. Let now $S_j$ be any element of $(G_l - I)$ not in the set $T_\alpha S_i T_\alpha^{-1}$. Then the sets $T_\alpha S_i T_\alpha^{-1}$ and $T_\alpha S_j T_\alpha^{-1}$ have no element in common. For from $T_\alpha S_i T_\alpha^{-1} = T_\beta S_j T_\beta^{-1}$ we should have $S_j = T_\beta^{-1} T_\alpha S_i T_\alpha^{-1} T_\beta$, which would mean that $S_j$ belongs to the set $T_\alpha S_i T_\alpha^{-1}$.

We see then that the $l-1$ elements $S_i (i = 2, \cdots, l)$ fall into sets each containing $a$ elements, and that no two of these sets have an element in common. Hence the truth of (4).

Suppose now that $G_M$ is any subgroup of $G_m$. Let

$$G_L = D(G_M, G_l),$$

that is, the group common to $G_M$ and $G_l$. We easily recognize that $G_L$ is invariant under $G_M$. The nature of the factor group $G_M/G_L$ is described by Lemmas 3 and 4.

LEMMA 3. *The factor group $G_M/G_L$ is isomorphic with a subgroup of $G_a$.*

The lemma follows immediately from a known theorem.*

LEMMA 4. *The factor group $G_M/G_L$ is isomorphic with $G_A$, a subgroup of $G_M$ having only the identity in common with $G_L$.*

If $G_L$ does not exhaust $G_M$ let us take as $G_a$ one of the $l$ isomorphic subgroups that has in common with $G_M$ at least one element other than the identity.

By (3), $G_m = G_l \cdot T_1 + \cdots + G_l \cdot T_a$; $G_a = (T_1 \cdots T_a)$. Let us suppose that $G_M$ is contained in the first $A$ complexes $G_l \cdot T_1 + \cdots + G_l \cdot T_A$ and that the $T$'s are so numbered that each complex actually contains at least one element of $G_M$. Comparison with the proof of Lemma 3 shows that the complex $(T_1 \cdots T_A)$ forms a subgroup $G_A$ of $G_M$ isomorphic with $G_M/G_L$, thus establishing the lemma.

From Lemma 4 follows without difficulty

LEMMA 5. *The complex $(G_M - G_L + L \cdot I)$ may be exhibited as the sum of $L$ non-overlapping conjugate groups each isomorphic with $G_M/G_L$.*

LEMMA 6. *Let $G_M$ be an invariant subgroup of $G_m$, such that $G_m/G_M$ is cyclic of order $\mu$. Then, either*

$$(5)\quad \begin{array}{ll} \text{(I)} & l = L \text{ and } a = \mu A; \text{ or} \\ \text{(II)} & l = \mu L \text{ and } a = A = 1. \end{array}$$

1. Assume (I) does not hold, and that $A > 1$.

Now the $L$ subgroups $G_A$ of $G_M$ are contained in $L$ of the $l$ subgroups $G_a$. Let $S$ (of $G_l$) transform one of these last groups into a $G_a$ having only the identity in common with the set of $L$ groups. Then the $G_A$ it contains is transformed into a group necessarily different from any of the $L$ subgroups $G_A$ contained in $G_M$. Hence, when $l > L$ and $A > 1$, $G_M$ cannot be invariant under $G_m$.

2. Suppose then that $l > L$, $A = 1$, but $a > A$.

Then

* Cf. Fricke, *Algebra*, 1924, vol. 1, p. 277

$$
\begin{aligned}
G_m = G_l G_a &= \sum_{i=1}^{1/L} \sum_{j=1}^{a} S_i G_L T_j \\
&= \sum \sum S_j G_M T_j = \sum \sum G_M S_i T_j \\
&= \sum_{i=0}^{\mu-1} G_M V^i .
\end{aligned} \tag{6}
$$

But $V = ST (T \neq I)$ is an element of $(G_m - G_l)$ and therefore of order $\leqq a$. But $\mu = la/L > a$; (6) is then impossible.

### 3. Derivation of the $\zeta$-relation

Dedekind* has shown how to derive the decomposition of a rational prime in a subfield of a Galois field when its decomposition in the larger field is assumed:

Let $K$ be a Galois field of degree $m$ and group $H_m$. Let $k$ be a subfield of $K$ of degree $\mu$ corresponding to a subgroup $H_\nu$ of order $\nu$, $m = \mu\nu$. Suppose that in $K$

$$
p = (\mathfrak{P}_1 \cdots \mathfrak{P}_e)^\lambda ; \ \mathfrak{P}_i \text{ of degree } g. \tag{7}
$$

Then in $k$,

$$
p = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_r^{a_r} ; \ \mathfrak{p}_i \text{ of degree } g_i' ;
$$

and

$$
\mathfrak{p}_i = (\mathfrak{P}_{i1} \cdots \mathfrak{P}_{i\rho_i})^{\lambda_i} ; \ \mathfrak{P}_{ij} \text{ of relative degree } g_i .
$$

The quantities $\tau$, $a_i$, $\rho_i$, $g_i$, $g_i'$, $\lambda_i$ must now be determined.

Let $H_\kappa$ be the "Zerlegungsgruppe"† of any prime ideal $\mathfrak{P}$ in the right member of (7). Let $H_m$ be decomposed with respect to the two subgroups $H_\nu$, $H_\kappa$:

$$
H_m = H_\nu \cdot V_1 \cdot H_\kappa + \cdots + H_\nu \cdot V_\tau \cdot H_\kappa ;
$$

the number of complexes is precisely $\tau$; no two complexes have an element in common; the number of elements in the $i$th complex is $\kappa\rho_i$, where $\sigma_i\rho_i = \nu$ and

$$
H_{\sigma_i} = D(V_i \cdot H_\kappa \cdot V_i^{-1}, H_\nu) .
$$

If $H_\lambda$ denote the "Trägheitsgruppe"‡ of the ideal $\mathfrak{P}$, then $\lambda_i$ is the order of the group

* *Zur Theorie der Ideale*, Göttinger Nachrichten, 1894, p. 272; or Fricke, *Algebra*, 1928, vol. 3, p. 186.

† Cf. Fricke, loc. cit., p. 171.

‡ Fricke, loc. cit.

$$H_{\lambda_i} = D(V_i \cdot H_\lambda \cdot V_i^{-1}, H_\tau).$$

Other relationships to be noted are

$$\kappa = g\lambda, \quad g = g_i g_i', \quad g_i\lambda_i = \sigma_i, \quad a_i\lambda_i = \lambda, \quad \sum\rho_i = e.$$

To apply the Dedekind theory, suppose $H_m \equiv G_m$ as defined above.

1. Let $H_\tau = G_l$:

$$G_m = G_l \cdot V_1 \cdot H_\kappa + \cdots + G_l \cdot V_\tau \cdot H_\kappa.$$

By Lemma 5, $H_\kappa = G_L \cdot G_A$, $G_L = D(H_\kappa, G_l)$; and since $H_\kappa$ satisfies all the assumptions made on $G_m$, we have $H_\lambda = G_{L'} \cdot G_{A'}$, where

$$G_{L'} = D(G_L, H_\lambda), \quad H_\lambda / G_{L'} = G_{A'}.$$

We have then

$$D(V_i \cdot H_\kappa \cdot V_i^{-1}, G_l) = V_i \cdot G_L \cdot V_i^{-1},$$

and

$$D(V_i \cdot H_\lambda \cdot V_i^{-1}, G_l) = V_i \cdot G_{L'} \cdot V_i^{-1}.$$

Therefore,

$$(8) \qquad \begin{aligned} &\sigma_i = L, \quad \rho_i = l/L, \quad \tau\rho_i = e, \quad \tau = eL/l; \\ &\lambda_i = L', \quad g_i = L/L', \quad g_i' = gL'/L, \quad a_i = \lambda/L'. \end{aligned}$$

2. Let $H_\tau = G_a$:

$$G_m = G_a \cdot V_1 \cdot H_\kappa + \cdots + G_a \cdot V_\tau \cdot H_\kappa.$$

We may suppose $G_a$ is that one of the $l$ conjugates that contains $G_A (A > 1$; when $A = 1$, any $G_a$ will serve the purpose). Evidently

$$D(V_i \cdot H_\kappa \cdot V_i^{-1}, G_a) = \begin{cases} G_{A'} \text{ for } i = 1, \\ G_1 \text{ for } i > 1; \end{cases}$$

whence

$$\sigma_1 = A, \quad \sigma_2 = \cdots = \sigma_\tau = 1;$$

$$\rho_1 = \frac{a}{A}, \quad \rho_2 = \cdots = \rho_\tau = a;$$

$$(9) \qquad D(V_i \cdot H_\lambda \cdot V_i^{-1}, G_a) = \begin{cases} G_A \text{ for } i = 1, \\ G_1 \text{ for } i > 1; \end{cases}$$

$$\lambda_1 = A', \quad \lambda_2 = \cdots = \lambda_\tau = 1; \quad g_1 = \frac{A}{A'}, \quad g_2 = \cdots = g_\tau = 1;$$

$$g_1' = \frac{gA'}{A}, \quad g_2' = \cdots = g_\tau' = g; \quad a_1 = \frac{\lambda}{A'}, \quad a_2 = \cdots = a_\tau = \lambda;$$

$$\frac{a}{A} + (\tau - 1)a = e; \quad \tau - 1 = \frac{e}{a} - \frac{1}{A}.$$

Now in $K$ we clearly have, using (7),

$$\zeta_K(s) = \prod_{\mathfrak{p}} \frac{1}{\left(1 - \frac{1}{p^{gs}}\right)^e}. \tag{10}$$

Let us fix our attention on a particular $G_\kappa$, and suppose it the Zerlegungsgruppe for some $\mathfrak{P}$, $\mathfrak{P}/p$. We then seek that factor of $\zeta_{k_a}(s)$ corresponding to $p$. By (8), the decomposition of $p$ in $k_a$ is

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_\tau)^{\lambda/L'}, \quad \tau = \frac{eL}{l};$$

each $\mathfrak{p}$ of degree $gL'/L$. Therefore the factor in question is

$$\frac{1}{\left(1 - \frac{1}{p^{gL's/L}}\right)^{eL/l}}. \tag{11}$$

As for the factor in $k_l$, we have, by (9), that $p$ decomposes in $k_l$ thus:

$$p = \mathfrak{p}_1^{\lambda/A'}(\mathfrak{p}_2 \cdots \mathfrak{p}_\tau)^\lambda, \quad \tau = 1 + \frac{e}{a} - \frac{1}{A};$$

degree of $\mathfrak{p}_1 = gA'/A$; degree of $\mathfrak{p}_i (i = 2, \cdots, \tau) = g$. We find then

$$\frac{1}{\left(1 - \frac{1}{p^{gA's/A}}\right)\left(1 - \frac{1}{p^{gs}}\right)^{e/a - 1/A}}. \tag{12}$$

To prove (1) we must then verify that the product of the expression (11) by the $a$th power of (12) is (using (10))

$$\frac{1}{\left(1 - \frac{1}{p^{gs}}\right)^e} \cdot \frac{1}{\left(1 - \frac{1}{p^s}\right)^a}.$$

We notice first that $G_\kappa$, necessarily a subgroup of $G_m$, has the same properties as $G_m$ (Lemma 5). Further it is known that $G_\lambda$ is invariant under $G_\kappa$ and that the factor group $G_\kappa/G_\lambda$ is cyclic.* Lemma 6 may then be applied,

* Fricke, loc. cit., p. 175.

and we find that either

(I) $L = L'$ and $A = gA'$; or

(II) $L = gL$ and $A = A' = 1$.

(I) We must show that

$$\left(1-\frac{1}{p^{ge}}\right)^{-eL/l}\left[\left(1-\frac{1}{p^{e}}\right)\left(1-\frac{1}{p^{ge}}\right)^{e/a-1/A}\right]^{-a}=\left(1-\frac{1}{p^{ge}}\right)^{-e}\left(1-\frac{1}{p^{e}}\right)^{-a};$$

which is correct, since

$$\frac{eL}{l}+e-\frac{a}{A}=e+\frac{eLA-la}{lA}=e+\frac{e\kappa-m}{lA}=e.$$

(II) Here we are to prove

$$\left(1-\frac{1}{p^{e}}\right)^{-eL/l}\left[\left(1-\frac{1}{p^{ge}}\right)^{e/a}\right]^{-a}=\left(1-\frac{1}{p^{ge}}\right)^{-e}\left(1-\frac{1}{p^{e}}\right)^{-a}.$$

This follows immediately from

$$\frac{eL}{l}=\frac{eLA}{la}\cdot a=\frac{e\kappa}{m}\cdot a=a.$$

The proof of (1) is then complete.

It is unnecessary to give the proof of (1a) since the Dedekind theory holds (with one obvious modification) for relative Galois fields.*

## 4. Proof of (2)

Let $\Omega$ be the fundamental form† of $K$, and let $\mathfrak{d}$ be the "differente" of $K$. Evidently

$$\mathfrak{d}=\prod_{\substack{U \text{ in } G_m \\ U\neq I}}(\Omega-U\Omega). \tag{13}$$

If $k$ be any subfield of $K$, we shall let $\mathfrak{d}(k)$ and $\mathfrak{D}(k)$ denote respectively the differente of $k$ and the differente of $K$ relative to $k$. As is well known‡

$$\mathfrak{d}=\mathfrak{d}(k)\mathfrak{D}(k). \tag{14}$$

If we suppose that $k$ corresponds to the subgroup $G_M$ of $G_m$, then

$$\mathfrak{D}(k)=\prod_{\substack{U \text{ in } G_M \\ U\neq I}}(\Omega-U\Omega). \tag{15}$$

* See, for example, Weber, *Algebra*, 1899, vol. 2, p. 657.

† Cf. Hilbert, loc. cit., p. 195.

‡ Hilbert, loc. cit., Satz 41.

Applying (15) we find that

$$\begin{aligned}
\mathfrak{D}(k_a) &= \prod(\Omega - U\Omega), \quad U \text{ in } (G_l - I),\\
\mathfrak{D}(k_l) &= \prod(\Omega - U\Omega), \quad U \text{ in } (G_a - I),\\
\mathfrak{D}(S_2k_l) &= \prod(\Omega - U\Omega), \quad U \text{ in } (S_2G_aS_2^{-1} - I),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\mathfrak{D}(S_lk_l) &= \prod(\Omega - U\Omega), \quad U \text{ in } (S_lG_aS_l^{-1} - I).
\end{aligned}$$

Multiplying together the corresponding members of these $l+1$ equations we get (using (13))

$$\mathfrak{d} = \mathfrak{D}(k_a)\mathfrak{D}(k_l)\mathfrak{D}(S_2k_l)\cdots\mathfrak{D}(S_lk_l); \tag{16}$$

or, by (14),

$$\mathfrak{d}^l = \mathfrak{d}(k_a)\mathfrak{d}(k_l)\mathfrak{d}(S_2k_l)\cdots\mathfrak{d}(S_lk_l).$$

But

$$d(k_l) = \mathfrak{d}(k_l)\mathfrak{d}(S_2k_l)\cdots\mathfrak{d}(S_lk_l),*$$

so that (16) may be written

$$\mathfrak{d}^l = \mathfrak{d}(k_a)d(k_l). \tag{17}$$

Now in a Galois field (of degree $m$) the discriminant is the $m$th power of the differente.† Hence, from (17) we infer

$$d = d(k_a)d^a(k_l). \tag{2}$$

It seems scarcely necessary to go into the proof of (2a); the remark made in the Introduction indicates how it may be readily derived from (2).

## 5. Abelian fields of type $(1, 1, \cdots, 1)$

We shall now briefly consider Abelian fields that are compounded of cyclic fields of equal prime degree. Assume then $K$ an Abelian field with group $G_{q^f}$ of type $(1, 1, \cdots,$ to $f$ units), $q$ a prime. $G_{q^f}$ contains $(q^f-1)/(q-1)$ subgroups of order $q^{f-1}$;‡ hence $K$ contains $(q^f-1)/(q-1)$ cyclic fields. If these fields be denoted by $k^1, \cdots, k^{r_f}$, where $r_f=(q^f-1)/(q-1)$, then we shall prove

$$\zeta^{r_f-1}\zeta_K = \zeta_{K^1}\cdots\zeta_{K^r}; \tag{18}$$

$$d(K) = d(k^1)\cdots d(k^{r_f}).\S \tag{19}$$

* Hilbert, loc. cit., Satz 37.

† Hilbert, loc. cit., Satz 68.

‡ Cf. Burnside, *Theory of Groups*, 1912, p. 60.

§ Cf. Pollaczek, loc. cit., p. 534.

(I) **Proof of (18).** Assume that in $K$, $p=(\mathfrak{P}_1 \cdots \mathfrak{P}_e)^\lambda$, each $\mathfrak{P}$ of degree $g$. We employ Dedekind's method (§3). Let $G_\kappa$ be the Zerlegungsgruppe of $\mathfrak{P}_1$ (and therefore of each of the remaining $\mathfrak{P}$'s). Since $G_\kappa/G_\lambda$ is cyclic, there are but two cases to consider: (1) $\kappa=\lambda$; (2) $\kappa=q\lambda$.

(1) $\kappa=\lambda=q^t$. Put $G_{q^f}=G_\nu \cdot V_1 \cdot G_\kappa + \cdots + G_\nu \cdot V_\tau \cdot G_\kappa$, where $G_\nu$ is one of the $r_f$ subgroups of $G_{q^f}$ of order $q$. (Corresponding to $G_\nu$ is then a field of degree $q^{f-1}$.)

Now

$$D(G_\kappa, G_\nu) = \begin{cases} G_q & \text{for } r_t \text{ choices of } G_\nu, \\ G_1 & \text{for } (r_f - r_t) \text{ choices of } G_\nu; \end{cases}$$

$$D(G_\lambda, G_\nu) = D(G_\kappa, G_\nu).$$

For $r_t$ choices of $G_\nu$, then, $\tau=e$, $g_i'=1$; for the remaining $(r_f-r_t)$ choices, $\tau=e/q$, $g_i'=1$. Typical factors of the associated $\zeta$-functions are then

$$\frac{1}{\left(1-\frac{1}{p^s}\right)^e} \text{ and } \frac{1}{\left(1-\frac{1}{p^s}\right)^{e/p}}, \text{ respectively.}$$

Since

$$er_t + e(r_f - r_t)/q = er_{f-1} + q^{f-1},$$

we may conclude that in this case $(\kappa=\lambda)$

$$\left(\prod \frac{1}{\left(1-\frac{1}{p^s}\right)^e}\right)^{r_t} \left(\prod \frac{1}{\left(1-\frac{1}{p^s}\right)^{e/q}}\right)^{r_f-r_t}$$

$$= \left(\prod \frac{1}{\left(1-\frac{1}{p^s}\right)^e}\right)^{r_{f-1}} \left(\prod \frac{1}{1-\frac{1}{p^s}}\right)^{q^{f-1}}, \tag{20}$$

the products extending only to primes $p$ for which $\kappa=\lambda$.

(2) $\kappa=q^t$, $\lambda=q^{t-1}$. Again put $G_{q^f}=G_\nu \cdot V_1 \cdot G_\kappa + \cdots + G_\nu \cdot V_1 \cdot G_\kappa$. Then $D(G_\kappa, G_\nu)=G_q$ for $r_t$ choices of $G_\nu$; and for these choices of $G_\nu$,

$$D(G_\lambda, G_\nu) = \begin{cases} G_q & r_{t-1} \text{ times;} \\ G_1 & \text{the remaining } q^{t-1} \text{ times.} \end{cases}$$

$D(G_\kappa, G_\nu)=G_1$ for $(r_f-r_t)$ choices of $G_\nu$; and for each such choice $D(G_\lambda, G_\nu)=G_1$.

Therefore,

$$\text{for } r_{t-1} \text{ choices of } G_\nu,\ \tau = e,\ g_i' = q;$$
$$\text{for } q^{t-1} \text{ choices of } G_\nu,\ \tau = e,\ g_i' = 1;$$

for $(r_f - r_i)$ choices of $G_\tau$, $\tau = e/q$, $g_i' = q$.

Factors of the corresponding $\zeta$'s are then

$$\frac{1}{\left(1-\frac{1}{p^{qs}}\right)^{e}}, \quad \frac{1}{\left(1-\frac{1}{p^{s}}\right)^{e}} \text{ and } \frac{1}{\left(1-\frac{1}{p^{qs}}\right)^{e/q}}, \text{ respectively.}$$

From this we may write, since $e \cdot q^{i-1} = q^{f-1}$, and

$$er_{i-1} + (r_f - r_i)e/q = er^{f-1},$$

$$\left(\prod\left(1-\frac{1}{p^{qs}}\right)^{-e}\right)^{r_{i-1}}\left(\prod\left(1-\frac{1}{p^{qs}}\right)^{-e/q}\right)^{r_f-r_i}\left(\prod\left(1-\frac{1}{p^{s}}\right)^{-e}\right)^{q^{i-1}}$$

$$= \left(\prod\left(1-\frac{1}{p^{qs}}\right)^{-e}\right)^{r_{f-1}}\left(\prod\left(1-\frac{1}{p^{s}}\right)^{-1}\right)^{q^{f-1}}, \tag{21}$$

the products extending only to primes $p$ for which $\kappa = q\lambda$.

From (20) and (21) follows immediately

$$\zeta^{q^{f-1}}\zeta_K = \prod_k \zeta_k, \tag{22}$$

the product extending over the $r_f$ fields of degree $q^{f-1}$. Making use of (22), (18) follows easily by induction on $f$.

(II) **Proof of (19).** We notice, to begin with, that

$$(G_{q^f} - I) = \sum (G_q - I),$$

the sum extending over the $r_f$ subgroups $G_q$ of $G_{q^f}$.

Then, if $\mathfrak{d}(k)$ and $\mathfrak{D}(k)$ have the same significance as in §4,

$$\mathfrak{d} = \mathfrak{d}(K) = \prod_k \mathfrak{D}(k), \tag{23}$$

the product extending over the $r_f$ subfields of $K$ of degree $q^{f-1}$. Equations (23) together with $\mathfrak{d} = \mathfrak{d}(k)\mathfrak{D}(k)$ imply

$$\mathfrak{d}^{r_f - 1} = \prod_1^{r_f} \mathfrak{d}(k).$$

Raising both members of this equality to the $q^{f-1}$th power, we find

$$d^{r_f-1}(K) = \pm \prod_1^{r_f} d(k), \quad k \text{ of degree } q^{f-1}. \tag{24}$$

Then, exactly as in proving (18), we may prove (19) by induction, making use of (24).

We may put (19) in a more general form that will be seen to include (24) as well. The proof consists of an easy induction so that we shall state the result only:

$$d^{\nu_{f,m}-\nu_{f,m-1}}(K) = \pm \prod_{1}^{\nu_{f,m}} d(k), \tag{25}$$

where the product extends over all the subfields of $K$ of degree $q^m$, and $\nu_{f,m}$ is the number of subgroups of $G_{q^f}$ of order $q^m$:

$$\nu_{f,m} = \frac{q^f - 1}{q - 1} \cdot \cdots \cdot \frac{q^{f-m+1} - 1}{q^m - 1}.$$

A similar generalization holds for (18):

$$\zeta^{\nu_{f,m}-\nu_{f-1,m-1}}\zeta_K^{\nu_{f-1,m-1}} = \prod_{1}^{\nu_{f,m}} \zeta_k, \tag{26}$$

the product being taken as in (25).

Finally, if we suppose $K$ relative Abelian, the base field being some $F$, while everything else remains unchanged, then we may rewrite (25) and (26) thus:

$$d^{\nu_{f,m}-\nu_{f-1,m-1}}(F)d^{\nu_{f-1,m-1}}(K) = \pm \prod d(k), \tag{25a}$$

$$\zeta_F^{\nu_{f,m}-\nu_{f-1,m-1}}\zeta_K^{\nu_{f-1,m-1}} = \prod \zeta_k, \tag{26a}$$

the products being taken as in (25).

## 6. Several illustrations and an application

It may be of interest at this point to give several examples illustrating the group $G_m$ defined in §1.

(I) We mention first the case treated by Artin.* The group $G_m$ may be defined as a group of transformations

$$x' = \alpha x + \beta,$$

$\alpha$ and $\beta$ being marks of a finite (Galois) field of degree $l = q^f$. The subgroup $G_l$ consists of the transformations

$$x' = x + \beta;$$

a particular $G_a$ is defined by $x' = \alpha x$, so that it is in this case cyclic.

(II) In this example, $G_l$ and $G_a$ are Abelian but $l$ may be divisible by more than one prime.

* Mathematische Annalen, vol. 89 (1923), p. 147.

Let $l=p_1^{e_1}\cdots p_r^{e_r}$ $(e_i\geqq 1)$; let $a$ be chosen so that $p_i\equiv 1$, mod $a$; let $u_i$ appertain to $a$, mod $p_i^{e_i}$; choose $u$ to satisfy $u\equiv u_i$, mod $p_i^{e_i}$; then we define $G_m$ as the group generated by $S$ and $T$, $S_i=I=T^a$, $TST^{-1}=S^u$. The group $G_l$ is generated by $S$; and $G_a$ may be taken as $T$ and its powers.

(III) For the third instance,* we use a subgroup $G_{72}$ of the Hessian group $G_{216}$.

The $G_{72}$ is generated by $S_1$, $S_2$, $T_1$, $T_2$, where

$$S_1^3 = S_2^3 = I, \quad S_2S_1 = S_1S_2;$$

$$T_1^4 = T_2^4 = I, \quad T_1^2 = T_2^2 \neq I, \quad T_2T_1 \neq T_1^3T_2;$$

$$T_1S_1T_1^{-1} = S_1S_2, \quad T_1S_2T_1^{-1} = S_1S_2^2,$$

$$T_2S_1T_2^{-1} = S_1^2S_2, \quad T_2S_2T_2^{-1} = S_1S_2.$$

$G_l$ consists of $S_1$, $S_2$, and products of powers of these elements; $G_l$ is evidently Abelian of type (1,1).

For a particular $G_a$ we take the group of order eight generated by $T_1$ and $T_2$. This group is not Abelian; it is indeed the quaternion group.

Returning to the first example, let $K$ be a Galois field with group $G_{la}$ as described $(l=q^f)$. Let $k_a$ be the field corresponding to $G_l$; it is evidently Abelian. Let $k_l$ be one of the $l$ conjugate fields of degree $l$. Suppose it possible to choose $m$ so that $0<m<f$, $q^m\equiv 1$, mod $a$. Then it is easily seen that any subfield of $K$ of degree $q^ma$ is Galoisian; furthermore its structure is exactly like that of $K$ if we merely replace $f$ by $m$.

To begin with, let us think of $K$ as relative Galoisian to $k_a$; the relative group is Abelian of type $(1, 1, \cdots, 1)$. By (26a), if $F\equiv k_a$,

$$\zeta_{k_a}^{rf,m-rf-1,m-1}\zeta_K^{rf-1,m-1} = \prod_1^{r_{f,m}}\zeta_k, \tag{27}$$

the product extending over all subfields of $K$ that are of degree $q^m$ relative to $k_a$ (that is, to $F$). Now, by (1),

$$\zeta^a\zeta_K = \zeta_{k_a}\zeta_{k_l}^a; \tag{28}$$

and $k_j'$ denoting one of the subfields of degree $j=q^m$ of some $k$ ($k$ as in (27)),

$$\zeta^a\zeta_k = \zeta_{k_a}\zeta_{k_j} \tag{29}$$

for each $k$. Substituting (28) and (29) in (27), we get

$$\zeta_{k_l}^{rf,m-rf-1,m-1}\zeta_{k_l}^{rf-1,m-1} = \prod_1^{r_{f,m}}\zeta_{k_j}, \tag{30}$$

* This was suggested by Professor Mitchell.

the product on the right extending over all fields of degree $j=q^m$ (each set of $j$ conjugate fields but one representative, namely, the one contained in $k_l$).

We may in exactly the same way derive a relation like (30) for discriminants:

$$d(k_l)^{\nu_{j-1,m-1}} = \pm \prod_1^{\nu_{j,m}} d(k_j). \tag{31}$$

## 7. A SECOND APPLICATION

We consider again the group $G_m$ of §1, but we shall now make an additional restriction: $a$ is some prime, $q$. Let us define a group $G_{lq^2}$ of order $lq^2$ as the direct product of $G_m$ and $H_q$, a group of order $q$. The group $G_{lq^2}$ then contains $q$ invariant subgroups isomorphic with $G_m$: If $G_q$ be generated by $T$, i.e., $G_q=\{T\}$, and $H_q=\{U\}$, then the $q$ subgroups are

$$G_l\cdot\{U^iT\} \qquad (i=1,\cdots,q).$$

Suppose now $K$ is an absolute Galois field with group $G_{lq^2}$ as defined above. To each $G_l\cdot\{U^iT\}$ corresponds a cyclic field $k_q^i$ of degree $q$; $K$ is a relative Galois field of relative degree $m$ with respect to $k_q^i$; the relative group is $G_l\cdot\{U^iT\}$, i.e. $G_m$. Accordingly, if in (1a) we let $F\equiv k_q^i$,

$$\zeta_{k_q^i}^q\zeta_K=\zeta_{K_q^i}\zeta_{K_l^i}^q \qquad (i=1,\cdots,q), \tag{32}$$

where $K_q^i$ is of relative degree $q$, and $K_l^i$ is of relative degree $l$. The field $K_q^i$ is then of absolute degree $q$; if we think of $K$ as an absolute Galois field, then, since $G_{lq^2}$ has but one subgroup of order $l$, evidently all the symbols $K_q^i$ denote the same field, $K_q$, say. Further $G_l$ is invariant under $G_{lq^2}$ and the factor group is Abelian of type (1, 1); therefore $K_q$ is Abelian with group of the same type. The $q$ fields $k_q^i$ are of course subfields of $K_q$; it must contain one other; namely, the field corresponding to the subgroup $G_l\times H_q$ ($\times$ means direct product); this field will be called $k_q$. By (18),

$$\zeta^q\zeta_{K_q}=\zeta_{k_q}\prod_{i=1}^{q}\zeta_{k_q^i}. \tag{33}$$

Multiplying together the $q$ equations (32), we find

$$\zeta_K\prod_{i=1}^{q}\zeta_{k_q^i}=\zeta_{K_q}\prod_{i=1}^{q}\zeta_{K_l^i}.$$

Applying (33), this may be written

$$\zeta^q\zeta_K=\zeta_{k_q}\prod_{i=1}^{q}\zeta_{K_l^i}. \tag{34}$$

For the discriminants we prove without any change in method

$$d^{q}(k_q^i)d(K) = \pm\, d(K_q)d^{q}(K_l^i), \tag{35}$$

and

$$d(K) = \pm\, d(k_q) \prod d(K_l^i), \tag{36}$$

corresponding to (32) and (34) respectively.

The special case mentioned in the Introduction does not fall under the above, but may be handled in much the same way. The group of $K$ (using the notation of §1) is $G_8$:

$$G_8 = \{S,T\}, \quad S^4 = I = T^2, \quad TST = S^{-1}.$$

To $\{T\}$ corresponds $k_4^1$; to $\{ST\}, k_4^2$; to $\{S\}, k_2$. Let $k_2^1, k_2^2$ correspond to $\{S^2, T\}$, $\{S^2, ST\}$, respectively. To the invariant subgroup $\{S^2\}$ corresponds an Abelian $A_4$, of type (1, 1), containing $k_2, k_2^1, k_2^2$. Now with respect to $k_2^i$ $(i=1, 2)$, $K$ is relative Abelian of type (1, 1): The three relative fields are $k_4^i$ (counted twice) and $A_4$. Then, by (26a), if $F \equiv k_2^i$, and $m=1$, $f=2$,

$$\zeta_{k_2^i}^2 \zeta_K = \zeta_{k_4^i}\zeta_{A_4},$$

whence

$$\zeta_{k_2^1}\zeta_{k_2^2}\zeta_K = \zeta_{A_4}\zeta_{k_4^1}\zeta_{k_4^2}.$$

But

$$\zeta^2\zeta_{A_4} = \zeta_{k_2}\zeta_{k_2^1}\zeta_{k_2^2};$$

therefore, finally,

$$\zeta^2\zeta_K = \zeta_{k_2}\zeta_{k_4^1}\zeta_{k_4^2}.$$

## 8. Comparison of regulators

After the fundamental Dirichlet-Dedekind expression, $h$, the class number of a field $k$, may be evaluated thus:

$$h\mu = \lim_{s=1} (s-1)\zeta_k(s), \tag{37}$$

$$\mu = \frac{2^{r_1+r_2}\pi^{r_2}R}{w|d^{1/2}|},$$

where

$r_1$ = number of real fields conjugate to $k$,
$r_2$ = number of pairs of imaginary fields conjugate to $k$,
$w$ = number of roots of unity in $k$,
$R$ = regulator in $k$.*

Let us return to the (absolute) Galois field $K$ with group $G_m$ as defined in §1. We shall assume in what follows that $K$ is real; we see then that

* See, for example, Hecke, *Die Theorie der Algebraischen Zahlen*, p. 156.

$$\mu(K) = \frac{2^{m-1}R(K)}{[d(K)]^{1/2}}, \quad \mu(k_l) = \frac{2^{l-1}R(k_l)}{[d(k_l)]^{1/2}}, \quad \mu(k_a) = \frac{2^{a-1}R(k_a)}{[d(k_a)]^{1/2}}.$$

Hence, by (1), (2), and (37)

$$\frac{h(K)}{h^a(k_l)h(k_a)} = \frac{R^a(k_l)R(k_a)}{R(K)}. \tag{38}$$

We shall now consider the quotient of regulators in the right member of (38).

Let $\epsilon_1, \cdots, \epsilon_{a-1}$ be a set of fundamental units in $k_a$; $\eta_1, \cdots, \eta_{l-1}$ a set of fundamental units in $k_l$. It will be convenient to employ the following abbreviations:

$$T_iE_j = \log|T_i\epsilon_j| \qquad (i = 1, \cdots, a; j = 1, \cdots, a-1),$$
$$S_iH_j = \log|S_i\eta_j| \qquad (i = 1, \cdots, l; j = 1, \cdots, l-1),$$

where, as in §2, $G_l = (S_1, \cdots, S_l)$, $G_a = (T_1, \cdots, T_a)$, $S_1 = T_1 = I$. We construct the set of $a$ distinct elements $T_iS_2T_i^{-1}(i = 1, \cdots, a)$. It will be convenient to denote the members of this set by $U_1, \cdots, U_a$.

We now consider the set of $(m-1)$ units $\epsilon_i$, $U_a\eta_j$ $(i = 1, \cdots, a-1;$ $j = 1, \cdots, l-1; \alpha = 1, \cdots, a)$. Our object is to prove that *at least when $l$ is a prime, they form a set of independent units in $K$.* Since much of the work goes through in the general case, we shall make no assumption about the nature of $l$ until it is necessary. The regulator of the units in question will be called $R_0$; it will be formed thus:

| | | |
|---|---|---|
| 1st row: $E_1 \cdots E_{a-1}\ U_1H_1 \cdots U_1H_{l-1} \cdots U_aH_{a-1}$ | | |
| 2d row: | $S_2$(1st row) | $l$ rows; |
| . . . . . | | |
| $l$th row: | $S_l$(1st row) | |
| $(l+1)$st row: | $T_2$(1st row) | |
| $(l+2)$d row: | $S_2((l+1)$st row) | $l$ rows; |
| . . . . . | | |
| $(2l)$th row: | $S_l((l+1)$st row) | |
| . . . . . | | |
| $((a-1)l+1)$st row: | $T_a$(1st row) | |
| . . . . . | | $(l-1)$ rows. |
| $(al-1)$st row: | $S_{l-1}[(a-1)l+1]$st row | |

A first simplification of this determinant can be effected by noticing that

(i) $SE_i = E_i$, (ii) $S_1H_j + \cdots + S_lH_j = 0$.

Let us add the 1st, 2d, $\cdots$, $(l-1)$st rows to the $l$th row; it will become

$$lE_1 \cdots lE_{a-1} 0\ 0 \cdots (al - a \text{ zeros}).$$

Similarly, if we add the $(l+1)$st, $\cdots$, $(2l-1)$st rows to the $(2l)$th row, it will become

$$lT_2E_1 \cdots lT_2E_{a-1} 0\ 0 \cdots (al - a \text{ zeros}).$$

And so on. It is clear that $R_0$ becomes

$$l^{a-1}\begin{vmatrix} E_1 & \cdots & E_{a-1} \\ T_2E_1 & \cdots & T_2E_{a-1} \\ \cdots & \cdots & \cdots \\ T_{a-1}E_1 & \cdots & T_{a-1}E_{a-1} \end{vmatrix} \cdot \Delta = \pm\, l^{a-1}R(k_a)\Delta, \tag{39}$$

where

$$\Delta = |M_{ij}|, \quad M_{ij} = \begin{Vmatrix} T_iU_jH_1 & \cdots & T_iU_jH_{l-1} \\ S_2T_iU_jH_1 & \cdots & \cdot \\ \cdots & \cdots & \cdots \\ S_{l-1}T_iU_jH_1 & \cdots & \cdot \end{Vmatrix}.$$

Now from the definition of $U_j$ it is clear that the set of matrices $M_{ij}$ ($i$ fixed; $j=1, \cdots, a$) is identical except for order with the set $M_{1j}$. Furthermore the set of permutations

$$\begin{pmatrix} M_{11} & M_{12} \cdots M_{1a} \\ M_{i1} & M_{i2} \cdots M_{ia} \end{pmatrix}$$

form a group isomorphic with $G_a$.

Let us write $M_{ij} = N_{ij}\overline{R}$,

$$\overline{R} = \| S_iH_j \| \qquad (i,j = 1, \cdots, l-1).$$

Then

$$\begin{aligned} \Delta &= |N_{ij}| \cdot |\overline{R}|^a \\ &= \pm\, |N_{ij}| \cdot R^a(k_l). \end{aligned} \tag{40}$$

The determinant $|N_{ij}|$ is a generalization of the class of determinants known as circulants. When $G_a$ is Abelian the determinant may be transformed into a product of determinants of the $(l-1)$st order; this is a direct generalization of a known result.*

Assuming then that $G_a$ is Abelian, let $\chi_\alpha(T)$ ($\alpha=1, \cdots, a$) represent the

* Burnside, Messenger of Mathematics, vol. 23, p. 112.

$a$ characters of the group.* Let

$$X_0 = |\chi_i(T_j)| \qquad (i,j = 1, \cdots, a);$$

then

$$|X_0|^2 = |\chi_i(T_j)| \cdot |\chi_i(T_j^{-1})| = |\sum_\alpha \chi_\alpha(T_i)\chi_\alpha(T_j^{-1})|$$
$$= |\sum_\alpha \chi_\alpha(T_iT_j^{-1})| = a^a.$$

Therefore

$$|X_0| = a^{a/2} \neq 0.$$

Let now $X = |\chi_i(T_j)E|$, where $E$ is the $(l-1)$-rowed unit matrix; clearly, $X = X_0^{l-1}$. Therefore

$$X \neq 0. \tag{41}$$

Returning to $|N_{ij}|$, we have

$$|N_{ij}| \cdot X = |N_{ij}| \cdot |\chi_i(T_j)E| = |\sum_\alpha \chi_i(T_\alpha)N_{j\alpha}|$$
$$= |\sum_\alpha \chi_i(T_\alpha T_j^{-1})N_{1\alpha}| = |\chi_i(T_j^{-1})\sum_\alpha \chi_i(T_\alpha)N_{1\alpha}|$$
$$= |\chi_i(T_j^{-1})E| \cdot \prod_i |\sum_\alpha \chi_i(T_\alpha)N_{1\alpha}|.$$

Therefore, by (41),

$$|N_{ij}| = \prod_i |\sum_\alpha \chi_i(T_\alpha)N_{1\alpha}|, \tag{42}$$

in absolute value.

Now exactly the same result obtains when $N_{1\alpha}$ is an ordinary (rather than a matric) quantity. We see then that a *determinant like* $|N_{ij}|$ *may be reduced to a product of determinants of order* $l-1$ *by treating its matric elements as if they were ordinary numbers.*

We suppose in the following that $G_m$ is a "congruence group, modulo $l$," that is, a group that may be defined as the set of transformations†

$$x' \equiv \alpha x + \beta \pmod{l}.$$

Let $r$ appertain to the index $a$, mod $l$; then, if we define a matrix $A$ of order $(l-1)$ by

$$A = \begin{Vmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ -1 & -1 & -1 & \cdot & \cdot & -1 \end{Vmatrix}, \tag{43}$$

* Weber, loc. cit., p. 49.

† Fricke, loc. cit., vol. 1, p. 443 ff.

it is easily seen that the set of matrices $N_{1\alpha}$ is now replaced by $A^{r^i}$ $(i=0, \cdots, a-1)$. It should be noticed that

$$A^l = E, \quad E + A + \cdots + A^{l-1} = 0. \tag{44}$$

Also, since $G_a$ is now cyclic, its characters may be expressed in terms of $\rho$, a primitive root of unity of index $a$.

From the above it readily follows that (42) becomes

$$|N_{ij}| = \prod_i |\sum_j \rho^{ij} A^{r^j}| . \tag{45}$$

The determinants in the right member may be evaluated by substituting for $A$ a primitive $l$th root of unity, $\zeta$. Now, $\sum_j \rho^{ij}\zeta^{r^j}$ is not zero; this follows from Kronecker's theorem on the irreducibility of the cyclotomic equation. We may then assert that $|N_{ij}| \neq 0$. Therefore, by (39) and (40), $R_0$ is not zero, so that the statement made at the beginning of this section is substantiated: *$\epsilon_i$, $U_a\eta_j$ form a set of independent units in $K$.*

Explicit results may be obtained very easily in two extreme cases: (I) $a=l-1$; (II) $a=2$.

(I) $a=l-1$. To evaluate the right member of (45) we note that $\sum_{j=0}^{a-1} \rho^{ij}\zeta^{r^j}$ is now the familiar Lagrange resolvent generally denoted by $(\rho^i, \zeta)$. For the resolvent we have

$$(\rho^0, \zeta) = -1 \text{ and } (\rho^i, \zeta)(\rho^{-1}, \zeta) = (-1)^i l \qquad (i = 1, \cdots, l-2).$$

Hence, in this case,

$$|N_{ij}| = l^{(l-1)(l-2)/2} \text{ in absolute value.}$$

Therefore

$$R_0 = l^{(l+1)(l-2)/2} R(k_a) R^a(k_l). \tag{46}$$

(II) $a=2$. $|N_{ij}|$ is here

$$\begin{vmatrix} A & A^{-1} \\ A^{-1} & A \end{vmatrix} = |A^2 - A^{-2}| = |A^{-2}| \cdot |A - E| \cdot |A + E| \cdot |A^2 + E| .$$

But, from (43), $|A| = 1$.

Secondly,

$$|A + E| \cdot |A^2 + E| \cdots |A^{l-1} + E| = 1,$$

whence $|A+E| = 1$, $|A^2+E| = 1$.

Thirdly,

$$|I - A| \cdot |I - A^2| \cdots |I - A^{l-1}| = |l| = l^{l-1};$$

therefore

$$|I - A| = l.$$

Finally, then,

$$R_0 = l^2 R(k_a) R^2(k_l).$$

Returning to (46), we shall now show that *the ratio of* $R_0$ *to* $R(K)$ *is a power* $(\geqq 0)$ *of* $l$.

Assume $R_0 > R(K)$; then there exist a positive integer $n$, and a unit $B$ of $K$, such that

$$\sum_1^{l-2} b_i E + \sum_1^{l-1} \sum_1^{l-1} b_{ij} S^i H_j = n \log |B|, \tag{47}$$

where

$$0 \leqq b_i < n, \quad 0 \leqq b_{ij} < n; \quad (b_1, \cdots, b_{l-2}, b_{11}, \cdots, b_{l-1,l-1}) = 1.$$

Putting (47) in exponential form, applying $T^u$, and reverting to the logarithmic form $(TST^{-1} = S^u)$,

$$\sum b_i T^u E_i + \sum \sum b_{ij} S^{ir^u} H_j = n \log |T^u B| \quad (u = 1, \cdots, l-2). \tag{48}$$

Adding (47) to the $(l-2)$ equations (48), we have

$$\sum_{i,j,u=1}^{l-1} b_{ij} S^{ir^u} H_j = n \log |B \cdot TB \cdots T^{l-2} B| = n \log |\alpha|, \tag{49}$$

$\alpha$ a unit in $k_l$.

Now

$$\begin{aligned} \sum_u S^{ir^u} H_j &= \sum_u S^{iu} H_j \\ &= \sum_u S^u H_j \text{ (since } 1 \leqq i \leqq l-1) \\ &= -H_j, \end{aligned}$$

so that (49) becomes

$$-\sum \sum b_{ij} H_j = n \log |\alpha|;$$

that is

$$-\sum_j \left( \sum b_{ij} \right) H_j = n \log |\alpha|.$$

But since the $\eta_j$ form a fundamental set in $k_l$, this last equation implies

$$n \mid \sum_i b_{ij}. \tag{50}$$

We return again to (47) and begin by applying $S^{-1}$:

$$\sum b_i E_i + \sum\sum b_{ij} S^{i-1} H_j = n \log |S^{-1}B| . \tag{51}$$

The double summation can be transformed thus:

$$\begin{aligned}\sum_{i,j=1}^{l-1} b_{ij} S^{i-1} H_j &= \sum_j b_{ij} H_j + \sum_j \sum_{i=2}^{l-1} b_{ij} S^{i-1} H_j \\ &= -\sum_i \sum_j b_{1j} S^i H_j + \sum_j \sum_{i=1}^{l-2} b_{i+1,j} S^i H_j \\ &= \sum_i \sum_j \bar{b}_{ij} S^i H_j,\end{aligned}$$

where

$$\begin{aligned}\bar{b}_{ij} &= b_{i+1,j} - b_{1j} \quad \text{for } 1 \leqq i \leqq l-2, \\ \bar{b}_{ij} &= -b_{1j} \qquad\quad \text{for } i = l-1.\end{aligned}$$

(51) may now be written as

$$\sum b_i E_i + \sum\sum \bar{b}_{ij} S^i H_j = n \log |\overline{B}| \qquad (\overline{B} = S^{-1}B). \tag{52}$$

But (52) is exactly like (47); therefore from (50)

$$n \mid \sum_i \bar{b}_{ij}.$$

(Note that in proving (50) no use is made of the auxiliary relations in (47).) This may be written, in terms of $b_{ij}$, in the following way:

$$n \mid \left( \sum_i b_{ij} - l b_{1j} \right). \tag{53}$$

In the same way, we may extend (53) to

$$n \mid \left( \sum_i b_{ij} - l b_{uj} \right) \qquad (u = 1, \cdots, l-1); \tag{54}$$

or, making use of (50),

$$n \mid l b_{ij} \qquad (i,j = 1, \cdots, l-1); \tag{55}$$

that is, $n \mid l \cdot (b_{11}, b_{12}, \cdots, b_{l-1,l-1})$.

Put $\sigma = (b_{11}, b_{12}, \cdots, b_{l-1,l-1})$, and assume

$$(n,\sigma) = \tau > 1.$$

Then (47) may be written

$$\sum b_i E_i = \tau \log |B_1| ,$$

or

$$\epsilon = B_1^{\tau}, \quad \epsilon = \epsilon_1^{b_1} \cdots \epsilon_{l-2}^{b_{l-2}}.$$

By one of the conditions in (47), $(\tau, b_1, \cdots, b_{l-2})=1$, so that $\epsilon$ is not the $\tau$th power of a unit in $k_a$. It remains to show that $\epsilon=B_1^{\tau}$ is impossible for any $B_1$ of $K$. Since $k_a$ is a maximal proper subfield of $K$, $\tau$ must be a multiple of $l$. Hence we need only consider the possibility of

$$\epsilon = B_1^{l}.$$

But as $K$ does not (under our assumptions) contain the $l$th roots of unity, this equality cannot hold.

We see then that $\tau$ is unity, and therefore

$$n \mid l. \tag{56}$$

Taking this in conjunction with (47), we see that our assertion on the ratio of $R_0$ to $R(K)$ is proved.

It is natural to enquire as to the number of independent relations like (47) actually existing in a field $K$. An exact answer apparently entails great difficulties; however it is easy to determine the maximum number of such relations. Making use of (50), this number is seen to be $(l-1)(l-2)$. Therefore, we have at once

$$\frac{R_0}{R(K)}\Bigg| l^{(l-2)(l-1)}.$$

(46) may then be written

$$hR(K) = l^{(l+1)(l-2)/2}R(k_a)R^{l-1}(k_l) \qquad (0 \leqq \lambda \leqq (l-1)(l-2)). \tag{57}$$

If (57) be compared with (36), we obtain the following class-number relationship $(a=l-1)$:

$$h(K) = l^{(l+1)(l-2)/2-\lambda}h(k_a)h^{l-1}(k_l). \tag{58}$$

It is of course rather obvious that the methods of this section may be applied to the other types of fields considered above. It does not however seem possible (using only these methods) to go as far as (57). On the other hand, Pollaczek, in the paper referred to above, obtains interesting results in the Abelian case (§5).

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

# A CANONICAL FORM OF GREEN'S PROJECTIVE ANALOGUE OF THE GAUSS DIFFERENTIAL EQUATIONS*

BY

V. G. GROVE

## 1. Introduction

The projective differential geometry of a surface in space of three dimensions may be founded upon Green's projective analogue of the Gauss differential equations.† In this paper we assume for the parametric curves any non-asymptotic system, and derive a canonical form for Green's equations. The method used is an adaptation of the method used by Lane in deriving a canonical form‡ for the defining differential equations of a conjugate net. By specializing the parametric curves to be conjugate, we may obtain the equivalent of Lane's canonical form. Again by assuming the parametric curves to be non-conjugate we obtain a canonical form for Green's differential equations§ defining a non-conjugate net. These canonical forms may be written with covariant derivatives.

## 2. An intermediate form of Green's equations

Let the homogeneous coördinates $y^{(1)}$, $y^{(2)}$, $y^{(3)}$, $y^{(4)}$ of a point $y$ of a non-degenerate surface $S_y$ be given as analytic functions of two independent variables $u, v$. Let also $z^{(1)}$, $z^{(2)}$, $z^{(3)}$, $z^{(4)}$ be the homogeneous coördinates of a point not in the tangent plane to $S_y$ at $y$. The four pairs of functions $(y, z)$ are solutions of a system of differential equations of the form¶

$$
\begin{aligned}
y_{uu} &= \alpha y_u + \beta y_v + py + Lz, \\
y_{uv} &= a y_u + b y_v + cy + Mz, \\
y_{vv} &= \gamma y_u + \delta y_v + qy + Nz, \\
z_u &= m y_u + s y_v + fy + Az, \\
z_v &= t y_u + n y_v + gy + Bz.
\end{aligned} \tag{1}
$$

Since system (1) is a symmetrical system, the formulas appear in pairs.

* Presented to the Society, December 31, 1928; received by the editors January 25, 1930.

† G. M. Green, *Memoir on the general theory of surfaces and rectilinear congruences*, these Transactions, vol. 20 (1919), p. 148. Hereafter referred to as Green, *Surfaces*.

‡ E. P. Lane, *Contributions to the theory of conjugate nets*, American Journal of Mathematics, vol. 49 (1927), pp. 565–576. Hereafter referred to as Lane, *Nets*.

§ G. M. Green, *Nets of space curves*, these Transactions, vol. 21 (1920), p. 207.

¶ Green, *Surfaces*, p. 149.

We shall give one of each pair; the other is obtained from the one given by the transformation scheme

$$S\begin{pmatrix} u, & \alpha, & \beta, & p, & L, & a, & c, & M, & m, & s, & f, & A \\ v, & \delta, & \gamma, & q, & N, & b, & c, & M, & n, & t, & g, & B \end{pmatrix},$$

any symbol being replaced by the one immediately above or below it.

The integrability conditions of system (1) are given by the following equations* and by the scheme $S$:

$$\begin{aligned} a_u - \alpha_v + ab + \beta\gamma + c &= Lt - Mm, \\ b_u - \beta_v + a\beta + b(b-\alpha) - \beta\delta - p &= Ln - Ms, \\ c_u - p_v + ap + c(b-\alpha) - \beta q &= Lg - Mf, \\ M_u - L_v + aL + M(b-\alpha) - \beta N &= LB - MA; \end{aligned} \tag{2a}$$

$$\begin{aligned} t_u - m_v + \alpha t + a(n-m) - s\gamma + g &= At - Bm, \\ g_u - f_v + pt + c(n-m) - sq &= Ag - Bf, \\ B_u - A_v + Lt + M(n-m) - sN &= 0. \end{aligned} \tag{2b}$$

Green has shown† from these conditions that

$$(a + \delta + A)_u = (b + \alpha + B)_v. \tag{3}$$

If we transform the independent variables by the transformation

$$\bar{u} = \phi(u,v),\ \bar{v} = \psi(u,v), \tag{4}$$

where

$$\phi_u = -\theta\phi_v,\quad \psi_v = -\omega\psi_u,\quad \theta\omega - 1 \neq 0, \tag{5}$$

we make the integral curves of

$$(\theta du - dv)(\omega dv - du) = 0$$

parametric. Transformation (4) changes system (1) into a system of like form, six of whose coefficients are given by $S$ and the formulas

$$\begin{aligned} \bar{\alpha} &= \frac{1}{\phi_v(\theta\omega - 1)^2}[\beta\omega^2 + 2b\omega + \delta - \theta(\alpha\omega^2 + 2a\omega + \gamma) \\ &\quad + \omega^2\theta_u + (\theta\omega - 2)\omega\theta_v] - \frac{\phi_{vv}}{\phi_v{}^2}, \\ \bar{\beta} &= \frac{\psi_u}{\phi_v{}^2(\theta\omega - 1)^2}[\alpha\omega^2 + 2a\omega + \gamma - \omega(\beta\omega^2 + 2b\omega + \delta) + \omega\omega_u + \omega_v], \end{aligned} \tag{6}$$

* Green, *Surfaces*, p. 150.

† Green, *Surfaces*, p. 152.

$$\bar{L} = \frac{1}{\phi_v^2(\theta\omega - 1)^2}[L\omega^2 + 2M\omega + N].$$

Since the asymptotic curves are not parametric, we may make $\bar{L}=0$ by taking

$$\omega = (d^{1/2} - M)/L, \quad d = M^2 - LN. \tag{7}$$

For this choice of $\omega$, the coefficient $\bar{\beta}$ assumes the form

$$\bar{\beta} = \frac{\psi_u \omega}{2\phi_v^2(\theta\omega - 1)d^{1/2}}(\omega L Q^{(u)} + N Q^{(v)}) \tag{8}$$

wherein $Q^{(u)}$ and $Q^{(v)}$ are defined by $S$ and the formula

$$Q^{(u)} = 2\left[\frac{2(aN - \gamma M)}{N} - \frac{\delta L - \beta N}{L}\right] + 2\frac{M}{L}\left[\frac{2(bL - \beta M)}{L} - \frac{\alpha N - \gamma L}{N}\right] - 2\left(\frac{M}{L}\right)_u - \frac{\partial}{\partial v}\log\frac{L}{N}. \tag{9}$$

From (8) and $S$ we find readily that

$$\bar{\beta}\bar{\gamma} = \frac{LNR}{16d(d^{1/2} - M)\phi_v\psi_u}, \tag{10}$$

wherein

$$R = LN(LQ^{(u)^2} - 2MQ^{(u)}Q^{(v)} + NQ^{(v)^2})/d. \tag{11}$$

Hence *the surface $S_y$ is ruled if and only if $R=0$.*

If we make the transformation

$$y = \lambda\bar{y} \tag{12}$$

on system (1), we obtain a new system of the same form with the following coefficients:

$$\begin{aligned} &\bar{\alpha} = \alpha - 2\lambda_u/\lambda, \ \bar{\beta} = \beta, \ \bar{p} = p + \alpha\lambda_u/\lambda + \beta\lambda_v/\lambda - \lambda_{uu}/\lambda, \ \bar{L} = L/\lambda, \\ &\bar{a} = a - \lambda_v/\lambda, \quad \bar{c} = c + a\lambda_u/\lambda + b\lambda_v/\lambda - \lambda_{uv}/\lambda, \qquad \bar{M} = M/\lambda. \end{aligned} \tag{13}$$

The remaining coefficients are either obtainable by $S$ or unnecessary for our purposes. The functions $d$, $Q^{(u)}$, $Q^{(v)}$, and $R$ are transformed by (12) into $\bar{d}$, etc. defined by the formulas

$$\bar{d} = d/\lambda^2, \quad \bar{Q}^{(u)} = Q^{(u)}, \quad \bar{Q}^{(v)} = Q^{(v)}, \quad \bar{R} = R/\lambda. \tag{14}$$

It follows therefore from (3) and (14) that there exists a function $F(u, v)$ such that

$$F_u = \alpha + b + A - 2R_u/R + (3/2)d_u/d,$$
$$F_v = \delta + a + B - 2R_v/R + (3/2)d_v/d. \tag{15}$$

Transformation (12) transforms the functions $F_u$, $F_v$ into

$$\overline{F}_u = F_u - 4\lambda_u/\lambda, \quad \overline{F}_v = F_v - 4\lambda_v/\lambda. \tag{16}$$

Hence if we choose $\lambda = e^{(1/4)F}$, we make $\overline{F}_u = \overline{F}_v = 0$. The resulting differential equations with coefficients (13) for this choice of $\lambda$ will be called the *intermediate form* of (1).

The transformation

$$\bar{u} = \phi(u), \ \bar{v} = \psi(v) \tag{17}$$

changes system (1) into another system whose coefficients are defined by $S$ and by

$$\begin{aligned} &\bar{\alpha} = (\alpha - \phi_{uu}/\phi_u)/\phi_u, \quad \bar{\beta} = \psi_v\beta/\phi_u, \quad \bar{p} = p/\phi_u^2, \quad \overline{L} = L/\phi_u^2, \\ &\bar{a} = a/\phi_v, \quad \bar{c} = c/(\phi_u\psi_v), \quad \overline{M} = M/(\phi_u\psi_v), \\ &\bar{m} = m, \quad \bar{s} = \psi_v s/\phi_u, \quad \bar{f} = f/\phi_u, \quad \overline{A} = A/\phi_u. \end{aligned} \tag{18}$$

The invariants $d$, $Q^{(u)}$, $Q^{(v)}$, and $R$ are transformed according to the formulas

$$d = d/(\phi_u\psi_v), \quad \bar{Q}^{(u)} = Q^{(u)}/\psi_v, \quad \bar{Q}^{(v)} = Q^{(v)}/\phi_u, \quad \overline{R} = R/(\phi_u\psi_v)^2.$$

Hence the conditions $F_u = F_v = 0$, characterizing the intermediate form, are unchanged by the transformation (17).

The polar reciprocal of the projective normal with respect to the quadric of Lie of the surface $S_y$ at $y$ intersects the asymptotic tangents in the points*

$$\bar{r} = \frac{\partial y}{\partial \bar{u}} - \frac{1}{2}\left[\bar{\alpha} + \frac{\partial}{\partial \bar{u}}\log(\bar{\beta}\bar{\gamma})\right]y, \quad \bar{s} = \frac{\partial y}{\partial \bar{v}} - \frac{1}{2}\left[\bar{\delta} + \frac{\partial}{\partial \bar{v}}\log(\bar{\beta}\bar{\gamma})\right]y, \tag{19}$$

wherein $\bar{u}$, $\bar{v}$ are the asymptotic parameters, and $\bar{\alpha}$, $\bar{\delta}$, $\bar{\beta}$, $\bar{\gamma}$ are defined by (6) and $S$. By means of (6) we may show that the reciprocal of the projective normal intersects the tangents to the parametric curves in the points

$$r = y_u - \tfrac{1}{4}F_u y,$$
$$s = y_v - \tfrac{1}{4}F_v y.$$

*The points $y_u$ and $y_v$ of the intermediate form are therefore characterized geometrically as the intersections of the reciprocal of the projective normal with the parametric tangents.*

* Green, *Surfaces*, p. 128.

### 3. The projective normal. The canonical form

The Darboux curves of the surface $S_y$ are defined by the differential equation

$$\bar{\beta}d\bar{u}^3 + \bar{\gamma}d\bar{v}^3 = 0, \tag{20}$$

wherein $\bar{\beta}$ and $\bar{\gamma}$ are defined by means of (6) and by $S$, and $\bar{u}$ and $\bar{v}$ are asymptotic parameters. In terms of the coefficients and variables of system (1) equation (20) may be written in the form

$$L^2(NQ^{(v)} - 2MQ^{(u)})du^3 - 3L^2NQ^{(u)}du^2dv - 3LN^2Q^{(v)}dudv^2 + N^2(LQ^{(u)} - 2MQ^{(v)})dv^3 = 0.$$

The asymptotic curves on $S_y$ are defined by the differential equation

$$Ldu^2 + 2Mdudv + Ndv^2 = 0.$$

Consider now the two forms

$$\begin{aligned} \phi_2 &= R(Ldu^2 + 2Mdudv + Ndv^2)/d, \\ \phi_3 &= R[L^2(NQ^{(v)} - 2MQ^{(u)})du^3 - 2L^2NQ^{(u)}du^2dv \\ &\quad - 3LN^2Q^{(v)}dudv^2 + N^2(LQ^{(u)} - 2MQ^{(v)})dv^3]/d^2. \end{aligned} \tag{21}$$

The ratio of the discriminant of $\phi_3$ to the cube of the discriminant of $\phi_2$ is a constant. Hence these forms are the forms $\phi_2$ and $\phi_3$ of Fubini.*

Since, for the intermediate form, $y_u$ and $y_v$ are the points in which the reciprocal of the projective normal intersects the parametric tangents, it follows that the projective normal joins $y$ to the point†

$$\zeta = Ny_{11} - 2My_{12} + Ly_{22},$$

wherein the $y_{ik}$ are the second covariant derivatives of the scalar $y$ with respect to the form $\phi_2$.

The Christoffel symbols of the second kind for the quadratic form $\phi_2$ are readily calculated to be defined by the formulas

$$\begin{aligned} 2d\begin{Bmatrix} 1 & 1 \\ & 1 \end{Bmatrix} &= 2LM(M/L)_u + (2M^2 - LN)[\log(LR/d)]_u - LM[\log(LR/d)]_v, \\ 2d\begin{Bmatrix} 2 & 2 \\ & 1 \end{Bmatrix} &= MN[\log(NR/d)]_v - LN[\log(LR/d)]_u, \\ 2d\begin{Bmatrix} 1 & 1 \\ & 2 \end{Bmatrix} &= L^2[\log(LR/d)]_v - LM[\log(LR/d)]_u - 2L^2(M/L)_u, \end{aligned} \tag{22}$$

* Fubini and Čech, *Geometria Proiettiva Differenziale*, Bologna, Zanichelli, 1926, pp. 84–87.

† Ibid., p. 87.

and by the substitution $S$.

In terms of the ordinary derivatives and the coefficients and variables of the intermediate form, the expression for $\zeta$ may be written

$$\zeta = \lambda y_u + \mu y_v + (\ )y - 2dz$$

wherein $\lambda$ and $\mu$ are defined by $S$ and the formula

$$2d\lambda = 2d(\alpha N - 2aM + \gamma L) + N(LN - 2M^2)[\log L/N]_u - LMN[\log L/N]_v - 2LMN(M/L)_u + 2LN^2(M/N)_v.$$

The point $Z$ defined by the expression

$$2dZ = -\lambda y_u - \mu y_v + 2dz$$

is evidently on the projective normal. If we make the transformation

$$z = Z + \tfrac{1}{2}\lambda y_u/d + \tfrac{1}{2}\mu y_v/d \tag{23}$$

on the intermediate form, we find that $(y, Z)$ are the solutions of differential equations of the form

$$\begin{aligned} y_{uu} &= \boldsymbol{\alpha} y_u + \boldsymbol{\beta} y_v + py + LZ, \\ y_{uv} &= \boldsymbol{a} y_u + \boldsymbol{b} y_v + cy + MZ, \\ y_{vv} &= \boldsymbol{\gamma} y_u + \boldsymbol{\delta} y_v + qy + NZ, \\ Z_u &= \bar{m} y_u + \bar{s} y_v + \bar{f} y + \bar{A} Z, \\ Z_v &= \bar{l} y_u + \bar{n} y_v + \bar{g} y + \bar{B} Z, \end{aligned} \tag{24}$$

wherein the new coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{a}$, $\boldsymbol{b}$ are defined by the formulas

$$\begin{aligned} 2d^2\boldsymbol{\alpha} &= \tfrac{1}{4}L^2MNQ^{(u)} - \tfrac{1}{4}L^2N^2Q^{(v)} + d(2M^2 - LN)[\log (R/N)]_u \\ &\quad - dLM[\log (R/L)]_v - M^2(2M^2 - LN)[\log (L/N)]_u \\ &\quad - LM(2M^2 - LN)[\log (L/N)]_v + 2LM^2N(M/N)_v - 2LM^3(M/L)_u, \\ 2d\boldsymbol{\beta} &= \tfrac{1}{4}L^2(LN - 2M^2)Q^{(u)} + \tfrac{1}{4}L^2MNQ^{(v)} - dLM[\log (R/N)]_u \\ &\quad - dL^2[\log (R/L)]_v + LM^2[\log (L/N)]_u \\ &\quad - L^2(LN - 2M^2)[\log (L/N)]_v + 2L^3N(M/L)_u - 2L^2MN(M/N)_v, \\ 2d\boldsymbol{a} &= \tfrac{1}{4}L^2NQ^{(u)} - \tfrac{1}{4}LMN^2Q^{(v)} + dMN[\log(R/N)]_u - dLN[\log (R/L)]_v \\ &\quad + MN(LN - 2M^2)[\log (L/N)]_u + LN(LN - 2M^2)[\log (L/N)]_v \\ &\quad - 2LM^2N(M/L)_u + 2LMN^2(M/N)_v, \end{aligned} \tag{25}$$

and by $S$. The coefficients $p$, $c$, $q$, $L$, $M$, $N$ are unchanged by (23). The remaining eight coefficients may be found in terms of the twelve thus defined by solving the proper equations of (2a) for the system (24).

If, as the point $y$ describes the curve $v=v(u)$, the line $yZ$ generates a developable, the function $v(u)$ must satisfy the following differential equation:

$$\bar{s}du^2 + (\bar{n} - \bar{m})dudv - \bar{t}dv^2 = 0. \tag{26}$$

If $Z+\lambda y$ is the point of contact of $yZ$ with the edge of regression of the developable, the function $\lambda$ must satisfy the quadratic equation

$$\lambda^2 + (\bar{m} + \bar{n})\lambda + \bar{m}\bar{n} - \bar{s}\bar{t} = 0. \tag{27}$$

From (27) we see that the harmonic conjugate of $y$ with respect to the focal points of $yZ$ is the point $z$ defined by

$$z = Z - \tfrac{1}{2}(\bar{m} + \bar{n})y.$$

If on system (24) we make the transformation

$$Z = z + \tfrac{1}{2}(\bar{m} + \bar{n})y \tag{28}$$

the coefficients $p$, $c$, $\bar{m}$, $\bar{f}$ are transformed according to the formulas

$$\begin{aligned} p &= p + \tfrac{1}{2}(\bar{m} + \bar{n}), \quad c = c + \tfrac{1}{2}(\bar{m} + \bar{n}), \\ m &= \tfrac{1}{2}(\bar{m} - \bar{n}), \quad f = \bar{f} + \tfrac{1}{2}\overline{A}(\bar{m} + \bar{n}) - \tfrac{1}{2}(\bar{m} + \bar{n})_u. \end{aligned} \tag{29}$$

The coefficients $q$, $n$ are obtained from (29) by $S$. The coefficients $L$, $M$, $N$, $\overline{A}$, $\overline{B}$ are unchanged by (28).

Since the projective lines of curvature defined by (26) form a conjugate system, it follows that

$$Lt + (n - m)M - sN = 0.$$

Therefore from the last of the integrability conditions (2b), we find that

$$\overline{A}_v - \overline{B}_u = 0.$$

If we now make the transformation

$$z = e^{\mu}z,$$

wherein

$$\mu_u = \overline{A}, \quad \mu_v = \overline{B}, \tag{30}$$

the new system of equations will have $A=B=0$. Hence *system* (1) *may be reduced to the following canonical form*:

$$\begin{aligned} y_{uu} &= \alpha y_u + \beta y_v + py + Lz, \\ y_{uv} &= ay_u + by_v + cy + Mz, \\ y_{vv} &= \gamma y_u + \delta y_v + qy + Nz, \\ z_u &= my_u + sy_v + fy, \\ z_v &= ty_u + ny_v + gy. \end{aligned} \tag{31}$$

The coefficients of this system are expressible in terms of the coefficients and variables of system (1) by equations (13), (16), (25), (29) and (30), and by solving the proper equations of (2a). *The system* (31) *is characterized analytically by the following conditions*:

$$\begin{aligned}
&\text{(a)}\quad F_u = F_v = 0,\\
&\text{(b)}\quad 2d(\alpha N - 2aM + \gamma L) + N(LN - 2M^2)[\log (L/N)]_u\\
&\qquad\qquad - LMN[\log (L/N)]_v - 2LMN(M/L)_u + 2LN^2(M/N)_v,\\
&\text{(c)}\quad m + n = 0,\\
&\text{(d)}\quad A = B = 0,
\end{aligned} \tag{32}$$

*and by the counterpart of* (b) *in* $S$.

Conditions (a) imply that the line $y_u\, y_v$ is the reciprocal of the projective normal; the conditions (b) and (a) imply that $yz$ is the projective normal; condition (c) implies that $z$ is the harmonic conjugate of $y$ with respect to the focal points on the projective normal; and the conditions (d) imply that the line of intersection of the tangent planes at $y$ and $z$ to $S_y$ and $S_z$ respectively is the line $z_u z_v$. *It is therefore evident geometrically that the system* (31) *characterized by* (32) *is unchanged in form by the transformation*

$$y = c\bar{y},\quad z = c'\bar{z},\quad \bar{u} = \phi(u,v),\quad \bar{v} = \psi(u,v),$$

*provided the transforms of* $L$ *and* $N$ *are not zero.*

We may write system (31) with covariant derivatives as follows:

$$\begin{aligned}
y_{11} &= \mathfrak{A}y_1 + \mathfrak{B}y_2 + py + Lz,\\
y_{12} &= \mathfrak{a}y_1 + \mathfrak{b}y_2 + cy + Mz,\\
y_{22} &= \mathfrak{C}y_1 + \mathfrak{D}y_2 + qy + Nz,\\
z_u &= my_1 + sy_2 + fy,\\
z_v &= ty_1 - my_2 + gy,
\end{aligned} \tag{33}$$

wherein the new coefficients $\mathfrak{A}$, $\mathfrak{B}$ etc. are defined by the following equations and by $S$:

$$\begin{aligned}
8d^2\mathfrak{A} &= L^2N(MQ^{(u)} - NQ^{(v)}),\\
8d^2\mathfrak{B} &= L^2MNQ^{(v)} + L^2(LN - 2M^2)Q^{(u)},\\
8d^2\mathfrak{a} &= L^2NQ^{(u)} - LMNQ^{(v)}.
\end{aligned} \tag{34}$$

## 4. Geometrical interpretation of the invariant coefficients of the canonical form

We shall now give geometrical interpretations of the invariant coefficients of system (31). We see readily that if $\beta = 0 (\gamma = 0)$ the curve $v$ = const.

($u$=const.) is a union curve of the congruence of projective normals. If $\beta=\gamma=0$ the axis congruence of the given parametric net coincides with the projective normal congruence. If $p=0$ ($q=0$) the tangent to $v$=const. ($u$=const.) on the surface $S_{y_u}(S_{y_v})$ generated by $y_u$ ($y_v$) intersects the line joining $y_v$ ($y_u$) to $z$.

The point $y_{uv}$ is the point of intersection of the tangents to $u$=const. on $S_{y_u}$ and the tangent to $v$=const. on $S_{y_v}$. Hence if $a=0$ ($b=0$) the point $y_{uv}$ lies in the plane determined by the tangent to $u$=const. ($v$=const.) and the projective normal. If the parametric net is non-conjugate, the projective normal and its reciprocal with respect to the quadric of Lie will be in Green's relation $R$ with respect to the given parametric net if and only if $a=b=0$. In case $M=0$, the parametric net is conjugate and the vanishing of $a$ and $b$ imply that the ray of the point $y$ with respect to the parametric net is the reciprocal of the projective normal.*

If $m=0$ the tangents to $u$=const. and $v$=const. on $S_z$ intersect respectively the tangents to $v$=const. and $u$=const. on $S_y$. If $s=0$ ($t=0$) the tangent to $v$=const. ($u$=const.) on $S_z$ intersects the tangent to $v$=const. ($u$=const.) on $S_y$. If $s=t=M=0$, the developables of the projective normal congruence intersect $S_y$ in the parametric curves. If $s=t=0$, $M\neq 0$, the projective normals pass through a point. If $f=0$ ($g=0$), the tangent to $v$=const. ($u$=const.) on $S_z$ intersects the reciprocal of the projective normal.

We may easily show that the developables of the reciprocal of the projective normal congruence correspond to the curves defined by the differential equation

$$(35)\qquad (cL-pM)du^2+(qL-pN)dudv+(qM-cN)dv^2=0.$$

These curves have been called the *reciprocal projective lines of curvature.*† The focal points of the reciprocal of the projective normal are defined by the expression $y_u+\lambda y_v$ where $\lambda$ satisfies the equation

$$(36)\qquad (cN-pM)\lambda^2+(pN-qL)\lambda+pM-cL=0.$$

The associate conjugate net of the parametric net is defined by the differential equation

$$(37)\qquad Ldu^2-Ndv^2=0.$$

The reciprocal projective lines of curvature therefore coincide with the associate conjugate curves if and only if

* Lane, *Nets*, p. 570.

† Lane, *Nets*, p. 570.

$$pN - qL = 0, \tag{38}$$

that is, if and only if the focal points of the reciprocal of the projective normal separate the points $y_u$ and $y_v$ harmonically.

Necessary and sufficient conditions that the ray of the point $y$ with respect to the parametric net coincide with the reciprocal of the projective normal are*

$$\mu = a - \gamma M/N = 0, \quad \lambda = b - \beta M/L = 0. \tag{39}$$

The condition that the ray tangents separate the parametric tangents harmonically is the vanishing of a certain invariant† $\mathfrak{D}$ defined by the formula

$$\begin{aligned} N\mathfrak{D} = {} & Np - Lq + L\mu^2 - N\lambda^2 + L\mu_v - N\lambda_u \\ & + (\beta N - L\delta)\mu - (\gamma L - \alpha N)\lambda + M(\mu_u - \lambda_v). \end{aligned} \tag{40}$$

Hence if conditions (38) and (39) are satisfied, the invariant $\mathfrak{D}$ vanishes. Therefore *if the associate conjugate net of the given net is the reciprocal projective lines of curvature, and if the ray of the given net is the reciprocal of the projective normal, the given net has equal point invariants of the first kind‡ and the ray tangents separate the tangents to the curves of the given net harmonically.* In case the given net is conjugate we have the theorem of Lane:§ If the associate of the given (conjugate) net is the reciprocal projective lines of curvature, and if the ray of the given net is the reciprocal of the projective normal, the given net has equal Laplace-Darboux invariants and is moreover harmonic.

## 5. A canonical form for the defining differential equations of a non-conjugate net

Let us consider a non-conjugate net. Since $M \neq 0$, we may eliminate $z$ from the first three of (31), obtaining the following system:

$$\begin{aligned} y_{uu} &= \mathfrak{A}y_{uv} + By_u + \mathfrak{C}y_v + \mathfrak{D}y, \\ y_{vv} &= \mathfrak{A}'y_{uv} + \mathfrak{B}'y_u + C'y_v + \mathfrak{D}'y, \end{aligned} \tag{41}$$

wherein the coefficients $\mathfrak{A}$, $B$, $\cdots$ are defined by the formulas

$$\begin{aligned} &\mathfrak{A} = L/M, \quad B = \alpha - a\mathfrak{A}, \quad \mathfrak{C} = \beta - b\mathfrak{A}, \quad \mathfrak{D} = p - c\mathfrak{A}, \\ &\mathfrak{A}' = N/M, \ \mathfrak{B}' = \gamma - a\mathfrak{A}', \ C' = \delta - b\mathfrak{A}', \ \mathfrak{D}' = q - c\mathfrak{A}'. \end{aligned} \tag{42}$$

* V. G. Grove, *Transformations of nets*, these Transactions, vol. 30 (1928), p. 489.

† V. G. Grove, *Nets with equal W invariants*, these Transactions, vol. 31 (1929), pp. 846–847.

‡ Ibid., p. 847.

§ Lane, *Nets*, pp. 570–571.

The coefficients written with German letters are invariants of the net.

The functions (15) in terms of the coefficients (42) are

$$F_u = f_u - 2\frac{R_u'}{R'} + \frac{3}{2}\frac{\partial}{\partial u}\log(1 - \mathfrak{A}\mathfrak{A}'),$$

$$F_v = f_v - 2\frac{R_v'}{R'} + \frac{3}{2}\frac{\partial}{\partial v}\log(1 - \mathfrak{A}\mathfrak{A}'),$$

wherein

$$R' = R/M$$

and the functions $f_u$ and $f_v$ are the same as the like named functions used by Green.*

The conditions (32b) may be written in terms of the coefficients (42) as follows:

$$\begin{aligned} 2(1 - \mathfrak{A}\mathfrak{A}')\boldsymbol{a} &= \mathfrak{A}'B + \mathfrak{A}\mathfrak{B}' + \mathfrak{A}_u' \\ &\quad + \frac{1}{2}\mathfrak{A}\mathfrak{A}'\left[\frac{\partial}{\partial v}\log(1 - \mathfrak{A}\mathfrak{A}') - \mathfrak{A}'\frac{\partial}{\partial u}\log(1 - \mathfrak{A}\mathfrak{A}')\right], \\ 2(1 - \mathfrak{A}\mathfrak{A}')\boldsymbol{b} &= \mathfrak{A}C' + \mathfrak{A}'\mathfrak{C} + \mathfrak{A}_v \\ &\quad + \frac{1}{2}\mathfrak{A}\mathfrak{A}'\left[\frac{\partial}{\partial u}\log(1 - \mathfrak{A}\mathfrak{A}') - \mathfrak{A}\frac{\partial}{\partial v}\log(1 - \mathfrak{A}\mathfrak{A}')\right]. \end{aligned} \tag{43}$$

If therefore the points $y_u$, $y_v$ are the points in which the reciprocal of the projective normal intersects the parametric tangents, the projective normal joins the point $y$ to the point $\zeta$ defined by

$$\zeta = y_{uv} - \boldsymbol{a}y_u - \boldsymbol{b}y_v. \tag{44}$$

In terms of covariant derivatives, system (41) assumes the form

$$\begin{aligned} y_{11} &= \mathfrak{A}y_{12} + \frac{\mathfrak{A}^2\mathfrak{A}'Q^{(u)}}{8(1 - \mathfrak{A}\mathfrak{A}')}y_1 + \frac{\mathfrak{A}^2(\mathfrak{A}'Q^{(v)} - 2Q^{(u)})}{8(1 - \mathfrak{A}\mathfrak{A}')}y_2 + \mathfrak{D}y, \\ y_{22} &= \mathfrak{A}'y_{12} + \frac{\mathfrak{A}'^2(\mathfrak{A}Q^{(u)} - 2Q^{(v)})}{8(1 - \mathfrak{A}\mathfrak{A}')}y_1 + \frac{\mathfrak{A}\mathfrak{A}'^2Q^{(v)}}{8(1 - \mathfrak{A}\mathfrak{A}')}y_2 + \mathfrak{D}y. \end{aligned} \tag{45}$$

The invariant coefficients of system (41) are readily interpreted. The line joining $y$ to $y_{uv}$ is the line in Green's relation $R$ to the reciprocal of the projective normal. Hence if $\mathfrak{C}=0$ ($\mathfrak{B}'=0$) the osculating plane of $v=$const. ($u=$const.) passes through the line $yy_{uv}$. If $\mathfrak{B}'=\mathfrak{C}=0$ the projective normal and its reciprocal are the axis and ray of the point $y$ with respect to the given

* G. M. Green, *Nets of space curves*, these Transactions, vol. 21 (1920), p. 212.

net and are therefore in Green's relation $R$ to that net. If $\mathfrak{D}=0$ ($\mathfrak{D}'=0$) the tangent to $v=$const. ($u=$const.) on the surface generated by $y_u$ ($y_v$) intersects the the tangent to $v=$const. ($u=$const.) on the surface generated by $y_v$ ($y_u$). If $\mathfrak{D}=\mathfrak{D}'=0$, the reciprocals of the projective normals lie in a fixed plane.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH.

# ON IRREDUCIBLE POLYNOMIALS IN SEVERAL VARIABLES WHICH BECOME REDUCIBLE WHEN THE VARIABLES ARE REPLACED BY POWERS OF THEMSELVES*

BY

ELI GOURIN

Let $Q(x_1, \cdots, x_s)$ be a polynomial in $x_1, \cdots, x_s$, irreducible in the field of all constants. There exist, in certain cases, integers $t_1, \cdots, t_s$ such that the polynomial $Q(x_1^{t_1}, \cdots, x_s^{t_s})$ is *reducible*.

The determination of the integers $t$ for which reducibility occurs is a problem which arose in an investigation of J. F. Ritt on the factorization of exponential forms.† The case of real interest is that in which $Q$ has at least three terms. We shall, in this introduction, limit ourselves to the discussion of the results for this case. The relatively simple case of two terms is treated in Part III.

For a fairly general, but not perfectly general, type of polynomial $Q$, consisting of more than two terms, Ritt proved that the sets $t$ break up into a finite number of classes, the sets of any one class being, from a certain point of view, equivalent.

In the present paper, we obtain information relative to the sets $t$ which, in certain respects, is final. Our results are embodied in the following theorems, in which each $t_i$ is understood to be a positive integer.

THEOREM I. *Let $Q(x_1, \cdots, x_s)$ be an absolutely irreducible polynomial, consisting of more than two terms. Suppose that at least one set $t_1, \cdots, t_s$ exists such that $Q(x_1^{t_1}, \cdots, x_s^{t_s})$ is reducible. Then there exists one and only one finite aggregate of sets*

$$(\alpha) \qquad t_{11}, \cdots, t_{1s}; \cdots; t_{n1}, \cdots, t_{ns}$$

*which fulfill the following conditions:*

(1) *For every $i$, $Q(x_1^{t_{i1}}, \cdots, x_s^{t_{is}})$ is reducible.*

(2) *If $Q(x_1^{t_1}, \cdots, x_s^{t_s})$ is reducible, there exists one and only one set $t_{j1}, \cdots, t_{js}$ such that each $t_k$ is an integral multiple of $t_{jk}$ and such that if $t_k = a_k t_{jk}$ $(k = 1, \cdots, s)$, then the irreducible factors of $Q(x_1^{t_1}, \cdots, x_s^{t_s})$ are found by replacing each $x_k$ by $x_k^{a_k}$ in the irreducible factors of $Q(x_1^{t_{j1}}, \cdots, x_s^{t_{js}})$.‡*

---

* Presented to the Society, February 23, 1929; received by the editors in February, 1930.

† *A factorization theory for functions* $\sum_{i=1}^{n} a_i e^{\alpha_i x}$, these Transactions, vol. 29 (1927), pp. 584–596.

‡ That is, if in any irreducible factor of $Q(x_1^{t_{j1}}, \cdots, x_s^{t_{js}})$ each $x_k$ is replaced by $x_k^{a_k}$ the resulting polynomial will be irreducible.

We shall call each of the $n$ sets of $(\alpha)$ a basic set.* With respect to such sets we obtain

THEOREM II. *Let $M=\max(M_1, \cdots, M_s)$, where $M_i$ is the degree of $Q(x_1, \cdots, x_s)$ in $x_i$. For any element $t_{ij}$ of any of the basic sets of $Q$ we have $t_{ij} \leqq M^2$.*

Furthermore, *the bound $M^2$ is the smallest possible bound. We construct polynomials $Q$ for which this bound is actually attained.*

The above results are stronger than those of Ritt in the following respects. Ritt deals with a type of polynomial $Q$ which he calls *primary* (see §1), and supposes that one of the terms of $Q$ is unity. In our case, $Q$ is any polynomial with more than two terms. But the chief advance of the present paper is the determination of the best upper bound, $M^2$, for the elements of the basic sets. The bound given by Ritt for the case with which he deals is $\delta^{s+4}$, where $\delta$ is the degree of $Q$. It may be remarked that our method of proof is essentially simpler than that of Ritt. His complicated first lemma is eliminated entirely.

In Part II, we define a set $t_1, \cdots, t_s$ as *minimal* if $Q(x_1^{t_1}, \cdots, x_s^{t_s})$ is reducible, but if no $Q(x_1^{\tau_1}, \cdots, x_s^{\tau_s})$ with each $\tau_i$ a submultiple of $t_i$, at least one $\tau_i$ a proper submultiple of its $t_i$, is reducible.

We prove, for a polynomial $Q$ consisting of more than two terms,

THEOREM III. *Those elements of a minimal set which are distinct from unity are equal to each other, and their common value is a prime number which does not exceed the greatest prime less than $M^2$.*

Furthermore we construct polynomials for which the upper bound given in Theorem III is actually attained.

## I. BASIC SETS

1. We shall say that two polynomials, neither identically zero, are equivalent if their ratio is a constant.

Let it be understood that no term of the polynomial $Q(x_1, \cdots, x_s)$ has a zero coefficient and that each $x$ is present in some term of $Q$ with an exponent greater than 0. Following Ritt, we shall say that $Q$ is *primary* in $x_i$ if the highest common factor of the exponents of $x_i$ in all the terms of $Q$ is unity. If $Q$ is primary in each of its variables, we say, simply, that $Q$ is *primary*.

Let $t_1, \cdots, t_s$ be positive integers. Consider the group $G$ of substitutions which replace the variables $x_1, \cdots, x_s$ by $\epsilon_1^{k_1}x_1, \cdots, \epsilon_s^{k_s}x_s$, respectively,

* Assuming the existence of these sets, it is easy to conclude that no infinite system of distinct sets satisfying the conditions (1) and (2) of the theorem can exist. See also §6.

where $\epsilon_i$ is a primitive $t_i$th root of unity and $k_i$ is an arbitrary integer. $G$ is of order $t_1 t_2 \cdots t_s$.

The $t_1 t_2 \cdots t_s$ transforms of a given polynomial by means of $G$ can be grouped into classes of equivalent polynomials. A set of transforms obtained by choosing one polynomial from each such class will be referred to as a *complete set of transforms.*

2. We shall prove the following lemma:

LEMMA I. *If* $Q(x_1, \cdots, x_s)$ *is irreducible and* $Q^{(t)} = Q(x_1^{t^1}, \cdots, x_s^{t^s})$ *is reducible, then the irreducible factors* $Q_1, \cdots, Q_N$ *of* $Q^{(t)}$ *form a complete set of transforms obtained from any one of them.*

Because $Q$ is irreducible, no monomial can be a factor of $Q$. Hence $Q^{(t)}$ can have no monomial factor, either. Thus $Q_1$ contains at least two terms.

It is obvious that every substitution of $G$ leaves $Q^{(t)}$ invariant. Consequently, any polynomial obtained from $Q_1$ by means of $G$ is, like $Q_1$, an irreducible factor of $Q^{(t)}$.

Let $Q_1, \cdots, Q_l$ be a complete set of transforms obtained from $Q_1$. The product $P = Q_1 \cdots Q_l$, for any substitution of $G$, goes over into a polynomial equivalent to $P$. Furthermore, as the $Q_i$'s are irreducible, and relatively prime, $P$ must be a factor of $Q^{(t)}$. Accordingly, let

$$Q^{(t)} = P \cdot R. \tag{1}$$

We shall prove that $R$ is a constant.

Consider any single variable, say $x_1$. Not every term of $P$ can contain $x_1$, else $P$, and therefore $Q^{(t)}$, would have a monomial factor. Hence the substitution of $G$ which replaces $x_1$ by $\epsilon_1 x_1$ and leaves all other variables unchanged must leave $P$ invariant. Consequently, $P$ is a rational function of $x_1^{t^1}$ and, similarly, of $x_i^{t^i}$, for every $i$. If, in (1), $R$ were not a constant, it would certainly be, as the quotient of $Q^{(t)}$ by $P$, a rational and, consequently, an integral rational function of every $x_i^{t_i}$. Thus $R$ must be a constant, else $Q$ would be reducible.

We may assume that $R$ is unity. Where this is not so at the start, it can be brought about by multiplying $Q_1$ by $1/R^{1/N}$.

Our lemma is thus proved.

COROLLARY. *Each* $Q_i$ *in the identity* $Q^{(t)} = Q_1 \cdots Q_N$ *involves every variable* $x_j$.

3. We assume now that the irreducible factor $Q_1$ of $Q^{(t)}$ is primary and

contains at least three terms. Because $Q_1$ is irreducible, it must contain at east one term independent of $x_1$. Suppose, then, that

$$(2)\quad Q_1 = A_1 x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_s^{\alpha_s} + A_2 x_1^{\beta_1} x_2^{\beta_2} \cdots x_s^{\beta_s} + \cdots + A_\mu x_1^{\mu_1} x_2^{\mu_2} \cdots x_s^{\mu_s} + A_\nu x_2^{\nu_2} \cdots x_s^{\nu_s}$$

where the $A$'s are constants and the last term does not contain $x_1$.

From (2) we find

$$(3)\quad Q_1 = x_2^{\nu_2} \cdots x_s^{\nu_s} U,$$

where

$$(4)\quad U = A_1 x_1^{\alpha_1} x_2^{\alpha_2-\nu_2} \cdots x_s^{\alpha_s-\nu_s} + A_2 x_1^{\beta_1} x_2^{\beta_2-\nu_2} \cdots x_s^{\beta_s-\nu_s} + \cdots + A_\mu x_1^{\mu_1} x_2^{\mu_2-\nu_2} \cdots x_s^{\mu_s-\nu_s} + A_\nu.$$

LEMMA II. *If $Q_1$ contains at least three terms and is primary, then, for at least one variable $x_i (i \neq 1)$ the exponents $\alpha_1, \beta_1, \cdots, \mu_1$ in $U$ will not be proportional to the corresponding exponents $\alpha_i - \nu_i, \beta_i - \nu_i, \cdots, \mu_i - \nu_i$.*

Suppose that this is not true. Then, since the exponents of $x_1$ in the various terms of $U$ are all non-negative integers, the exponents of any other variable $x_j$ must be either all non-positive or all non-negative. Let $D_j$ be the greatest common factor of the exponents of $x_j$ in $U$ (if they are non-positive, we shall understand by $D_j$ their common negative divisor of maximum absolute value).

The exponents of $x_j$ in $U$ will, accordingly, be of the form $a_{1j}D_j, \cdots, a_{\mu j}D_j$, where the $a$'s are relatively prime non-negative integers. When $j=1, D=1$, because $U$ has the same exponents for $x_1$ as $Q_1$. From this fact and the assumed proportionality of the exponents it follows immediately that, for any $j$, the set of integers $a_{1j}, \cdots, a_{\mu j}$ is identical with the set $\alpha_1, \cdots, \mu_1$. Consequently, $U$ is a polynomial in the product $z = x_1 x_2^{D_2} \cdots x_s^{D_s}$. Then

$$(5)\quad U = C_0 z^n + C_1 z^{n-1} + \cdots + C_{n+1},$$

with constant $C$'s and $n > 1$.

If $z_1, \cdots, z_n$ are the zeros of this latter polynomial, we find, as a consequence from (3) and (5),

$$(6)\quad Q_1 = C_0 x_2^{\nu_2} \cdots x_s^{\nu_s} (z - z_1) \cdots (z - z_n).$$

The product $z = x_1 x_2^{D_2} \cdots x_s^{D_s}$ must involve negative powers; otherwise, by (6), $Q$ would be reducible. Fixing our ideas, let us assume that $D_2, D_3, \cdots, D_k$, but no other $D$'s, are negative. Because the exponents of $Q_1$ are non-negative, we conclude from (6) that in the expressions

$$(7) \qquad d_2 = \nu_2 + D_2 n, \quad d_3 = \nu_3 + D_3 n, \cdots, d_k = \nu_k + D_k n$$

the $d$'s are non-negative integers.

From (6) and (7) we have that

$$(8) \quad Q_1 = C_0 x_2^{d_2} \cdots x_k^{d_k} x_{k+1}^{\nu_{k+1}} \cdots x_s^{\nu_s} \left[x_2^{-D_2} \cdots x_k^{-D_k}(z - z_1)\right] \cdots \left[x_2^{-D_2} \cdots x_k^{-D_k}(z - z_n)\right].$$

Each of the factors $x_2^{-D_2} \cdots x_k^{-D_k}(z - z_i)$ is a binomial in which the variables involved have positive exponents. Because $n$ is greater than unity, there are at least two such factors. This result is, however, an absurdity, since $Q_1$ is irreducible.

Thus in $U$, for some $x_i$, the exponents are not proportional to those of $x_1$.

4. We shall prove the following lemma:

LEMMA III. *If an irreducible factor $Q_1$ of $Q^{(t)} = Q(x_1^{t_1}, \cdots, x_s^{t_s})$ consists of at least three terms, and is primary, then each $t_j$, where $j$ is among the numbers $1, 2, \cdots, s$, satisfies the relation $t_j \leqq M^2$.*

It follows from Lemma I that, if $Q^{(t)} = Q_1 \cdots Q_N$, and $Q_1$ consists of more than two terms and is primary, then each of the $Q_i$'s contains, likewise, more than two terms and is primary.

We conclude further from Lemma II, and from the expression of $Q_1$ as given in (2), that there exists a subscript $i$ such that at least two numbers in the set $\alpha_1, \beta_1, \cdots, \mu_1$ are not proportional to two corresponding numbers in the set $\alpha_i - \nu_i, \beta_i - \nu_i, \cdots, \mu_i - \nu_i$. Changing, if necessary, the subscripts of the variables as well as the notation of the exponents in (2), we may assume that the numbers $\alpha_1, \beta_1$ are not proportional to $\alpha_2 - \nu_2, \beta_2 - \nu_2$ and that the determinant $\alpha_1(\beta_2 - \nu_2) - \beta_1(\alpha_2 - \nu_2)$ is positive.

Let $n_1$ be the degree of $Q_1$ in $x_1$ and $n_2$ the degree in $x_2$. Let $m_1$ and $m_2$ be, respectively, the degrees of $Q$ in the same variables. Because $Q_1$ is primary, there will be no two equivalent polynomials among the $t_1$ polynomials $Q_1(\epsilon_1^{k_1} x_1, x_2, \cdots, x_s)$ $(k_1 = 0, 1, \cdots, t_1 - 1)$.

Suppose, indeed, that, for distinct values $p$ and $q$ of $k_1$, the corresponding polynomials $Q_{1p}$ and $Q_{1q}$ satisfy a relation

$$Q_{1p} = cQ_{1q},$$

where $c$ is a constant. It is easy to see from (2) that $c$ must be unity and that

or

$$\epsilon_1^{p\alpha_1} = \epsilon_1^{q\alpha_1}, \cdots, \epsilon_1^{p\mu_1} = \epsilon_1^{q\mu_1},$$

$$\frac{\alpha_1(p - q)}{t_1} = I_\alpha, \cdots, \frac{\mu_1(p - q)}{t_1} = I_\mu,$$

where the $I$'s are integers. Because $\alpha_1, \cdots, \mu_1$ are relatively prime, $p-q$ must be divisible by $t_1$, which implies that $p=q$.

It follows therefore from Lemma I that $N$ is at least equal to $t_1$. By comparing the degrees in $x_1$ of both sides of the identity

$$Q^{(t)} = Q_1 \cdot Q_2 \cdots Q_N, \tag{9}$$

we find that

$$t_1 n_1 \leqq t_1 m_1, \quad \text{or} \quad n_1 \leqq m_1. \tag{10}$$

Similarly, we find that

$$t_2 n_2 \leqq t_2 m_2, \quad \text{or} \quad n_2 \leqq m_2. \tag{11}$$

Let us now denote by $Q_{uv}$ any of the polynomials which are obtained from $Q_1$ by placing $\epsilon_1^u x_1$ for $x_1$ and $\epsilon_2^v x_2$ for $x_2$, where $u$ is among the numbers $0, 1, \cdots, t_1-1$ and $v$ among the numbers $0, 1, \cdots, t_2-1$.

These $t_1 t_2$ polynomials may form several sets of equivalent polynomials. Suppose, therefore, that, for two pairs of integers $(u_1, v_1)$ and $(u_2, v_2)$, the corresponding polynomials $Q_{u_1 v_1}$ and $Q_{u_2 v_2}$ are equivalent. It follows then from (2) by considering the first two and the last terms, that the following relations must be satisfied:

$$\epsilon_1^{\alpha_1 u_1} \epsilon_2^{\alpha_2 v_1} = c\epsilon_1^{\alpha_1 u_2} \epsilon_2^{\alpha_2 v_2}, \qquad \epsilon_1^{\beta_1 u_1} \epsilon_2^{\beta_2 v_1} = c\epsilon_1^{\beta_1 u_2} \epsilon_2^{\beta_2 v_2}, \qquad \epsilon_2^{\nu_2 v_1} = c\epsilon_2^{\nu_2 v_2},$$

where $c$ is a constant. This implies that

$$\frac{\alpha_1(u_2 - u_1)}{t_1} + \frac{\alpha_2(v_2 - v_1)}{t_2} = I_1 + e, \tag{12}$$

$$\frac{\beta_1(u_2 - u_1)}{t_1} + \frac{\beta_2(v_2 - v_1)}{t_2} = I_2 + e, \tag{13}$$

$$\frac{\nu_2(v_2 - v_1)}{t_2} = I_3 + e, \tag{14}$$

where the $I$'s are integers and $e$ is a rational fraction.

Subtracting (14) from (12) and (13), we find that

$$\frac{\alpha_1(u_2 - u_1)}{t_1} + \frac{(\alpha_2 - \nu_2)(v_2 - v_1)}{t_2} = I_4,$$

$$\frac{\beta_1(u_2 - u_1)}{t_1} + \frac{(\beta_2 - \nu_2)(v_2 - v_1)}{t_2} = I_5,$$

and, solving for $(u_2-u_1)/t_1$ and $(v_2-v_1)/t_2$, we have

$$\frac{(u_2-u_1)\Delta}{t_1} = I_6, \tag{15}$$

$$\frac{(v_2-v_1)\Delta}{t_2} = I_7, \tag{16}$$

where $I_4, \cdots, I_7$ are integers and $\Delta$ is the determinant $\alpha_1(\beta_2-\nu_2) - \beta_1(\alpha_2-\nu_2)$.

Let $d_i (i=1, 2)$ be the highest common factor of $\Delta$ and $t_i (i=1, 2)$. Then

$$\Delta = d_1\delta_1 = d_2\delta_2, \quad t_1 = d_1\tau_1, \quad t_2 = d_2\tau_2,$$

where $\delta_1, \delta_2, \tau_1, \tau_2$ are positive integers.

Substituting these expressions in (15) and (16) we find that

$$\frac{u_2-u_1}{\tau_2}\delta_1 = I_6, \tag{17}$$

$$\frac{v_2-v_1}{\tau_2}\delta_2 = I_7. \tag{18}$$

Since the pairs $(\delta_1, \tau_1)$ and $(\delta_2, \tau_2)$ are relatively prime, the relations (17) and (18) can be satisfied only if $(u_2-u_1)$ is divisible by $\tau_1$ and $(v_2-v_1)$ by $\tau_2$.

It follows immediately that these relations will surely not be satisfied if $u_1$ and $u_2$ are both among the numbers $0, 1, \cdots, t_1-1$, and $v_1$ and $v_2$ both among the numbers $0, 1, \cdots, \tau_2-1$. Thus we can obtain from $Q_1$, by means of the substitutions of $G$, at least $t_1\tau_2$ polynomials no two of which are equivalent. This implies that $N \geqq t_1\tau_2$. Comparing the degrees in $x_2$ of both sides of the identity (9), we find that

$$t_1\tau_2 n_2 \leqq t_2 m_2, \tag{19}$$

or, since $\tau_2 = t_2/d_2$,

$$t_1 \leqq \frac{m_2 d_2}{n_2} \leqq \frac{m_2\Delta}{n_2}. \tag{20}$$

Now $\Delta$ stands for the expression $\alpha_1(\beta_2-\nu_2)-\beta_1(\alpha_2-\nu_2)$, which is positive. If, therefore, both $\beta_2-\nu_2$ and $\alpha_2-\nu_2$ are non-negative, then $\Delta \leqq \alpha_1(\beta_2-\nu_2) \leqq \alpha_1 n_2$. If they are both non-positive, then $\Delta \leqq \beta_1(\nu_2-\alpha_2) \leqq \beta_1 n_2$. Finally, if $\beta_2-\nu_2$ is positive and $\alpha_2-\nu_2$ is negative, then

$$\Delta = \alpha_1(\beta_2-\nu_2) + \beta_1(\nu_2-\alpha_2) \leqq n_1(\beta_2-\nu_2) + n_1(\nu_2-\alpha_2) \leqq n_1(\beta_2-\alpha_2) \leqq n_1 n_2.$$

We find, thus, that in all cases $\Delta \leqq n_1 n_2$. Hence, it follows from (20) that

$$t_1 \leqq m_2 n_1, \tag{21}$$

and, from (10), that

$$t_1 \leqq m_2 m_1 \leqq M^2. \tag{22}$$

This result we obtained by considering a particular variable $x_1$. We can apply the same reasoning to any other variable $x_i$ and prove that, in general,

$$t_i \leqq M^2. \tag{23}$$

Lemma III is proved.

5. We assume, next, that the irreducible polynomial $Q(x_1, \cdots, x_s)$ consists of at least three terms and prove

LEMMA IV. *If $Q$ consists of at least three terms and if the irreducible factor $Q_1$ of $Q^{(t)}$ is primary, then $Q_1$ consists, likewise, of at least three terms.*

Suppose that $Q_1$ contains just two terms.

Because $Q_1$ is irreducible, a given variable $x_i$ can be present only in one of the two terms. Further, since $Q_1$ is primary, it can involve only first powers of the variables. Let us assume, therefore, since we are free to interchange the subscripts, that

$$Q_1 = ax_1x_2 \cdots x_k + bx_{k+1} \cdots x_s, \tag{24}$$

where $a$ and $b$ are constants.

We denote by $q$ the ratio $x_{k+1} \cdots x_s/(x_1x_2 \cdots x_k)$, so that $Q_1 = ax_1 \cdots x_k \cdot (1+c_1q)$, where $c_1 = b/a$.

It follows from Lemma I that, if $Q^{(t)} = Q_1Q_2 \cdots Q_N$, then each $Q_i$ will be of the form $Q_i = d_iax_1 \cdots x_k(1+c_iq)$, where $d_i$ and $c_i$ are constants and $d_1 = 1$.

We find, in this way, the following expression for $Q^{(t)}$:

$$Q^{(t)} = Da^N(x_1x_2 \cdots x_k)^N(1 + B_1q^{\alpha_1} + B_2q^{\alpha_2} + \cdots + B_nq^N), \tag{25}$$

where $D = d_1d_2 \cdots d_N$ and the last factor is the product of the $N$ binomials $(1+c_iq)$.

Since $Q^{(t)}$ does not change when $x_i (i = 1, \cdots, k)$ is replaced by $\epsilon_ix_i$, it is easy to see that $N$ and each $\alpha_j$ must be divisible by $t_i$.

Similarly we conclude that $N$ and each of the $\alpha$'s must be divisible by $t_j$, where $j$ is among the numbers $k+1, \cdots, s$. Consequently, $N = \nu T$ and $\alpha_j = \gamma_jT$, where $T$ is the least common multiple of the numbers $t_1, \cdots, t_s$, and $\nu$ and $\gamma_j$ are positive integers. Denoting $q^T$ by $q_1$, we find from (25)

$$Q^{(t)} = Da^N(x_1 \cdots x_k)^{\nu T}(1 + B_1q_1^{\gamma_1} + B_2q_1^{\gamma_2} + \cdots + B_nq_1^{\nu}). \tag{26}$$

The exponent $\nu$ is greater than unity. Otherwise $Q^{(t)}$ and, consequently, $Q$ would contain only two terms. Then, the last factor of $Q^{(t)}$ in (26), as a polynomial in one variable of degree higher than unity, is reducible. Let $e_1, \cdots, e_\nu$ be the zeros of this polynomial.

We find from (26), substituting for $q_1$ its expression in the $x$'s,

$$Q^{(t)} = B_n D a^N [(x_{k+1} \cdots x_s)^\tau - e_1(x_1 \cdots x_k)^\tau] \cdots [(x_{k+1} \cdots x_s)^\tau - e_\nu(x_1 \cdots x_k)^\tau], \tag{27}$$

which contradicts the assumption that $Q$ is irreducible. Hence $Q_1$, and, consequently, each $Q_i$, consists of at least three terms.

6. An immediate consequence of the last two lemmas is the following result: Given an irreducible polynomial $Q(x_1, \cdots, x_s)$ consisting of more than two terms, there can exist only a finite number of sets

$$(\beta) \qquad t_{11}, \cdots, t_{1s}; \cdots; t_{n1}, \cdots, t_{ns}$$

such that the irreducible factors of $Q(x_1^{t_{i1}}, \cdots, x_s^{t_{is}})$ $(i=1, \cdots, n)$ are primary.

We assume that all sets of such nature are present among the $n$ sets of $(\beta)$ and, furthermore, that no two sets of $(\beta)$ are identical.

We shall prove that if $Q$ is primary, then the sets $(\beta)$ are the basic sets of $Q$ referred to in the introduction. Let, indeed, $t_1, \cdots, t_s$ be a set of exponents for which $Q^{(t)}$ is reducible, and let $P(x_1, \cdots, x_s)$ be one of the irreducible factors of $Q^{(t)}$. If $\alpha_j$ is the greatest common factor of the exponents of $x_j$ in $P$, so that

$$P(x_1, \cdots, x_s) = Q_1(x_1^{\alpha_1}, \cdots, x_s^{\alpha_s})$$

where $Q(x_1, \cdots, x_s)$ is primary, we conclude from Lemma I that

$$Q^{(t)} = Q_1(x_1^{\alpha_1}, \cdots, x_s^{\alpha_s}) \cdots Q_N(x_1^{\alpha_1}, \cdots, x_s^{\alpha_s}). \tag{28}$$

This implies that $Q^{(t)}$ is an integral rational function in $x_j^{\alpha_j}(j=1, \cdots, s)$.

Because $Q$ is primary, each $t_j$ must be divisible by the corresponding $\alpha_j$. If, therefore, $t_j=\alpha_j\tau_j$, we find from (28) that

$$Q(x_1^{\tau_1}, \cdots, x_s^{\tau_s}) = Q_1(x_1, \cdots, x_s) \cdots Q_N(x_1, \cdots, x_s) \tag{29}$$

where the $Q_i$'s are primary.

Consequently, the set $\tau_1, \cdots, \tau_s$ must be one of the sets $(\beta)$.

Hence, given any set of $t$'s for which $Q^{(t)}$ is reducible, there always exists, among the sets $(\beta)$, one and only one set, say $t_{k1}, \cdots, t_{ks}$, such that, for every $i$, $t_i=\alpha_i t_{ki}$ and such that the irreducible factors of $Q^{(t)}$ are obtained by replacing in the irreducible factors of $Q(x_1^{t_{k1}}, \cdots, x_s^{t_{ks}})$ each $x_i$ by $x_i^{\alpha_i}$.

Furthermore, any other aggregate $(\gamma)$ of sets having the two properties just stated must be identical with the aggregate $(\beta)$. We shall prove, first, that any set of $(\beta)$ belongs to $(\gamma)$. Otherwise, we must conclude that there exists in $(\beta)$ a set, say $t_{i1}, \cdots, t_{is}$, whose elements are integral multiples of the corresponding elements of some set of $(\gamma)$, one element, at least, a proper multiple. Let $\tau_1, \cdots, \tau_s$ be the elements of this latter set. Each $\tau_j$ is, in its turn, a multiple of the corresponding element in some set, say $t_{k1}, \cdots, t_{ks}$, of $(\beta)$, distinct from the set $t_{i1}, \cdots, t_{is}$. The elements of the two chosen sets of $(\beta)$ satisfy, therefore, relations of the form $t_{ij}=a_j t_{kj}$, at least one of the $a$'s being greater than unity. Moreover, the irreducible factors of $Q(x_1^{t_{i1}}, \cdots, x_s^{t_{is}})$ can be found by replacing in the irreducible factors of $Q(x_1^{t_{k1}}, \cdots, x_s^{t_{ks}})$ each $x_j$ by $x^{a_j}$, which is an absurdity. Hence each set of $(\beta)$ belongs to $(\gamma)$. Similarly it can be shown that each set of $(\gamma)$ belongs to $(\beta)$. The two aggregates are therefore identical.

The sets $(\beta)$ are thus the basic sets of $Q$.

7. Suppose, next, that the polynomial $Q$ is not primary. Let $\lambda_j$ be the greatest common factor of the exponents of $x_j$ in $Q$. We have, then, that

$$Q(x_1, \cdots, x_s) = Q'(x_1^{\lambda_1}, \cdots, x_s^{\lambda_s})$$

where $Q'(x_1, \cdots, x_s)$ is primary.

Consider the basic sets $(\beta)$ of $Q'$. In any of these sets, say $t_{i1}, \cdots, t_{is}$, we replace each $t_{ij}$ by $\tau_{ij}$, where $\tau_{ij}=t_{ij}/d_{ij}$ and $d_{ij}$ is the greatest common factor of $t_{ij}$ and $\lambda_j$. We obtain in this way, say, $n$ sets

$$(\beta') \qquad \tau_{11}, \cdots, \tau_{1s}; \cdots; \tau_{n1}, \cdots, \tau_{ns}.$$

We shall prove that the sets $(\beta')$ have the qualities necessary for them to be a system of basic sets of $Q$.

We verify, first, that, for any $i(i=1, \cdots, n)$, the polynomial $Q(x_1^{\tau_{i1}}, \cdots, x_s^{\tau_{is}})$ is reducible. We have, indeed, denoting $\lambda_j/d_{ij}$ by $\lambda_{ij}$, that

$$Q(x_1^{\tau_{i1}}, \cdots, x_s^{\tau_{is}}) = Q'(x_1^{\lambda_{i1}t_{i1}}, \cdots, x_s^{\lambda_{is}t_{is}}).$$

The latter polynomial is obviously reducible.

Further, if, for a given set of $t$'s, $Q^{(t)}$ is reducible, we find from (28) that

$$(30) \quad Q^{(t)} = Q'(x_1^{t_1\lambda_1}, \cdots, x_s^{t_s\lambda_s}) = Q_1(x_1^{\alpha_1}, \cdots, x_s^{\alpha_s}) \cdots Q_N(x_1^{\alpha_1}, \cdots, x_s^{\alpha_s}).$$

This implies that, for every $j$, $t_j\lambda_j$ must be divisible by $\alpha_j$. Hence, if $t_j\lambda_j=\alpha_j\tau_j$ we find from (30) that

$$(31) \qquad Q'(x^{\tau_1}, \cdots, x^{\tau_s}) = Q_1(x_1, \cdots, x_s) \cdots Q_N(x_1, \cdots, x_s)$$

where all the $Q_i$'s are primary.

Consequently, the $\tau$'s form one of the sets $(\beta)$, say the set $t_{k1}, \cdots, t_{ks}$. For every $j$ we have, therefore,

$$t_j = \frac{\alpha_j t_{kj}}{\lambda_j} = \frac{\alpha_j \tau_{kj}}{\lambda_{kj}}.$$

Because $\tau_{kj}$ and $\lambda_{kj}$ are relatively prime, $\alpha_j$ must be divisible by $\lambda_{kj}$. Accordingly, let $\alpha_j = \delta_j \lambda_{kj}$, so that $t_j = \delta_j \tau_{kj}$.

Replacing now in (30) each $x_j^{\delta_j}$ by $x_j$, we find that

$$(32) \quad Q(x_1^{\tau_{k1}}, \cdots, x_s^{\tau_{ks}}) = Q_1(x_1^{\lambda_{k1}}, \cdots, x_s^{\lambda_{ks}}) \cdots Q_N(x_1^{\lambda_{k1}}, \cdots, x_s^{\lambda_{ks}}).$$

Hence, given a set of $t$'s for which $Q^{(t)}$ is reducible, there exists always in $(\beta')$ one and only one set, say $\tau_{k1}, \cdots, \tau_{ks}$, such that, for every $j$, $t_j = \delta_j \tau_{kj}$, and such that the irreducible factors of $Q^{(t)}$ are obtained by replacing in the irreducible factors of $Q(x_1^{\tau_{k1}}, \cdots, x_s^{\tau_{ks}})$ each $x_j$ by $x_j \delta_j$.

We know already that there can exist no other system of sets having the same two properties.

The sets $(\beta')$ are, therefore, the basic sets of $Q$.

8. The results obtained in §§6 and 7 verify thus Theorem I, announced in the introduction. We conclude, further, on the basis of Lemma III, that Theorem II of the introduction is likewise true for a primary polynomial $Q$. If $Q$ is not primary, the upper bound for the elements of its basic sets, as we have seen, cannot exceed the corresponding bound for the basic sets of a primary polynomial whose degrees in the individual variables never exceed the degrees of $Q$ in the same variables. Hence, if $t$ is any element of any of the basic sets of $Q$, we have, a fortiori, that $t \leqq M^2$. Theorem II is, therefore, true for any polynomial $Q$.

9. We shall show now that the bound of Theorem II is the smallest possible bound. We shall construct a class of polynomials for which $t$ actually reaches the value $M^2$.

Consider the polynomial

$$(33) \qquad P_0(x_1, \cdots, x_s) = 1 + x_1 x_2 \cdots x_s + x_1^m,$$

where $m$ is a positive integer greater than unity.

We shall prove that $P$ *is irreducible for any* $m$. For, if $P_0 = P_1 \cdot P_2$, it is clear that $x_1$ must be present in both polynomials $P_1$ and $P_2$; otherwise, the coefficient of the last term in $P_0$ would be distinct from unity. This implies, however, that the polynomial

$$R(x_1, x_2) = 1 + x_1 x_2 + x_1^m$$

obtained from $P$ by replacing $x_3, x_4, \cdots, x_s$ by unity, is also reducible. If,

therefore, $R=R_1 \cdot R_2$, we conclude, for the same reason as before, that $x_1$ must be present in both $R_1$ and $R_2$. Because $R$ is linear in $x_2$, this variable can be present only in one of the factors, say $R_1$. Hence.

$$1+x_1x_2+x_1^m=R_1(x_1,x_2)R_2(x_1). \tag{34}$$

Any root $\alpha$ of the equation $R_2=0$ must be distinct from zero. Replacing in (34) $x_1$ by $\alpha$, we find that

$$1 + x_2\alpha + \alpha^m = 0,$$

where $x_2$ is arbitrary, which is, of course, an absurdity. Consequently, $P_0$ is irreducible.

Consider the $m$ polynomials $P_0, P_1, \cdots, P_{m^2-1}$ obtained from $P_0$ by substituting $\epsilon_1^{k_1}x_1$ for $x_1$, where $\epsilon_1$ is a primitive $m^2$th root of unity and $k_1$ ranges over the numbers $0, 1, \cdots, m^2-1$. Because $P_0$ is a primary polynomial, there will be no pair of equivalent polynomials in the set thus obtained. On the other hand, it is easy to verify that any polynomial obtained from $P_0$ by replacing $x_1$ by $\epsilon_1^{k_1}x_1$, where $\epsilon_1$ and $k_1$ have the same meanings as above, and by replacing, further, any other variable $x_j$ by $\epsilon_j^{k_j}x_j$, where $\epsilon_j$ is a primitive $m$th root of unity and $k_j=0, \cdots, m-1$, is identical with one of the polynomials $P_i$. The index $i$ is the smallest non-negative number satisfying the congruence

$$y \equiv [k_1 + m(k_2 + k_3 + \cdots + k_s)] \pmod{m^2}.$$

The product $\Pi(x_1, \cdots, x_s)$ of the $m$ polynomials $P_i$, therefore, remains invariant for any of the substitutions of the group $G$ described above and, consequently, is an integral rational function in $x_1^{m^2}, x_2^m, \cdots, x_s^m$.

Consider now the polynomial $\Pi_1(x_1, \cdots, x_s)$ obtained from $\Pi(x_1, \cdots, x_s)$ by replacing $x_1^{m^2}$ by $x_1$ and $x_j^m$ by $x_j (j=2, \cdots, s)$. The polynomial $\Pi_1$ is irreducible. For, if

$$\Pi_1 = A(x_1, \cdots, x_s)B(x_1, \cdots, x_s),$$

then

$$\Pi = A(x_1^{m^2}, x_2^m, \cdots, x_s^m)B(x_1^{m^2}, x_2^m, \cdots, x_s^m),$$

a relation which is absurd, considering that the group $G$ is transitive with respect to any of the polynomials $P_i$ the product of which is $\Pi$.

Now, the degree of $\Pi$ in $x_1$ is $m^3$, and in any of the remaining variables is $m^2$. Consequently, the degrees of $\Pi_1$ in the same variables are all equal to $m=M$. Furthermore, replacing in $\Pi_1$ the variable $x_1$ by $x_1^{M^2}$ and every other variable $x_j$ by $x_j^M$, we find that

$$\Pi_1(x_1^{M^2}, x_2^M, \cdots, x_s^M) = P_0 \cdot P_1 \cdots P_{M^2-1}$$

where the $P_i$'s are primary.

Hence $\Pi_1$ has, at least, one basic set in which one of the elements, namely $t_1$, actually attains the value $M^2$. This proves that the bound of Theorem II is the smallest possible bound.

## II. Minimal sets

10. From the definition of a minimal set, as given in the introduction, it follows easily that every minimal set of an irreducible polynomial consisting of more than two terms is necessarily a basic set. For its elements must be integral multiples of the corresponding elements of a particular basic set; no element, however, can be a proper multiple. Hence the minimal sets are finite in number.

11. Let $t_1, \cdots, t_s$ be the elements of a minimal set belonging to a primary irreducible polynomial $Q$. Then, in the identity

$$Q^{(t)} = Q_1 \cdots Q_N,$$

each $Q_i$ is primary. It is obvious, further, that not every $t_i$ is unity. Fixing our ideas, let us assume that $t_1 = t_2 = \cdots = t_{j-1} = 1$ and that the remaining $t_j, \cdots, t_s$ are all greater than unity. Consider the $t_j$ polynomials $Q_1(x_1, \cdots, \epsilon_j^{k_j} x_j, \cdots, x_s)$, $k_j = 0, \cdots, t_j - 1$. We have shown above (§4) that, because $Q$ is primary, no two among these polynomials are equivalent. Denoting their product by $P$, we find that

$$Q(x_1, \cdots, x_{j-1}, x_j^{t_j}, \cdots, x_s^{t_s}) = P \cdot R, \tag{35}$$

where $P$ and $R$ are integral rational functions in $x_j^{t_j}$. However, $R$ must be a constant; else the polynomial

$$Q(x_1, \cdots, x_{j-1}, x_j, x_{j+1}^{t_{j+1}}, \cdots, x_s^{t_s})$$

would be reducible and the set $1, \cdots, 1, t_j, \cdots, t_s$ would not be a minima set. Hence $N = t_j$. Similarly we find that $N = t_k$, for $k = j, j+1, \cdots, s$. Consequently, $t_j = t_{j+1} = \cdots = t_s = t$.

12. Suppose, now, that $t = q \cdot r$, where $q$ and $r$ are positive integers and $q$ is greater than unity. The product of the $t$ polynomials $Q_1(x_1, \cdots, \epsilon_j^{k_j} x_j, \cdots, x_s)$ can then be decomposed into $q$ products $P_0, P_1, \cdots, P_{q-1}$, where the factors of $P_i$ are the $r$ polynomials obtained by assigning to $k_j$ the $r$ values $i, i+q, \cdots, i+(r-1)q$. The polynomial $P_i$, therefore, will be invariant for any substitution which replaces $x_j$ by $\eta^k x_j$, where $\eta = e^{2\pi i/r}$ and $k = 0, 1, \cdots, r-1$. Consequently, $P_i$ is a rational integral function

in $x_j^r$. We have, then, that

$$Q(x_1, \cdots, x_{j-1}, x_j^t, \cdots, x_s^t) = P_0(x_1, \cdots, x_j^r, \cdots, x^t) \cdots P_{q-1}(x_1, \cdots, x_j^r, \cdots, x_s)$$

or

$$Q(x_1, \cdots, x_{j-1}, x_j^q, \cdots, x_s^t) = P_0(x_1, \cdots, x_j, \cdots, x_s) \cdots P_{q-1}(x_1, \cdots, x_j, \cdots, x_s)$$

which is a contradiction, unless $r=1$ and $q=t$.

Consequently $t$ is a prime number. Hence, if $Q$ is primary, those elements in a minimal set which are distinct from unity are all equal to one and the same prime number.

13. We shall consider now the case when $Q$ is not primary. Referring to §7, we shall recall that the basic sets of $Q$ are obtained from the basic sets $(\beta)$ of a certain primary polynomial by substituting for any element in each set $(\beta)$ a properly chosen factor of the element. It is clear, further, from the way these substitutions were defined, that all minimal sets of $Q$ will be found by effecting the indicated substitutions only in the minimal sets of $(\beta)$. The elements in each of the latter sets being equal either to unity or to one and the same prime number, it is obvious that the minimal sets of $Q$ will, thus, necessarily have the same structure.

14. Let $p$ be the common value of those elements in a minimal set which are distinct from unity. Because a minimal set is a basic set, and $p$ is a prime number, it follows from Theorem II that $p \leqq P$, where $P$ is the largest prime less than $M^2$.

The results obtained in §§12, 13, and 14 of this section verify thus Theorem III of the introduction.

15. We shall show now that the bound $P$, as defined in §14, is the smallest possible bound.

Consider, indeed, the polynomial

$$(36) \qquad Q_0(x_1, \cdots, x_s) = 1 + x_1^{m-q} x_2 \cdot x_3 \cdots x_s + x_1^{(m-q)m-P}(x_2 \cdot x_3 \cdots x_s)^m,$$

where $m$ is any integer greater than unity, $P$ is the largest prime less than $m^2$, and $q$ is the positive integer satisfying the inequalities

$$m^2 - (q+1)m < P < m^2 - qm.$$

We prove, first, that the polynomial $Q_0$ is primary in all its variables. This is obvious with regard to the variables $x_2, x_3, \cdots, x_s$. If, on the other hand, $d$ is the greatest common divisor of the exponents in $x_1$, then $d$ must

be a factor of both $m-q$ and $P$; consequently, if $d$ is distinct from unity, $P$ must be a factor of $m-q$. As $m-q<P$, we have $d=1$.*

We show, next, that $Q_0$ is irreducible. For, if $Q_0=Q_1\cdot Q_2$, it is clear that $x_1$ must be present in both $Q_1$ and $Q_2$. The polynomial

$$R(x_1, x_2) = 1 + x_1^{m-q}x_2 + x_1^{(m-q)m-P}x_2^m,$$

obtained from $Q$ by replacing each of the variables $x_3, \cdots, x_s$ by unity, is also reducible. Consider the equation $R(x_1, x_2)=0$. Treating $x_2$ as a function of $x_1$, we find, by means of Newton's polygon, that for the neighborhood of $x_1=0$

$$x_2 = \epsilon x_1^{-(m^2-qm-P)/m} + \cdots$$

where $\epsilon$ is an $m$th root of unity. Consequently, if $x_1$ describes a small circle in the complex plane about $x_1=0$, the $m$ branches of $x_2$ will be permuted in a single cycle. These $m$ branches thus hang together. Hence $R(x_1, x_2)$ cannot be reducible, which implies that $Q_0$ is irreducible.

Consider the $P$ polynomials $Q_0, Q_1, \cdots, Q_{P-1}$, obtained from $Q_0$ by substituting $\eta^{k_1}x_1$ for $x_1$, where $\eta$ is a $P$th root of unity and $k=0, 1, \cdots, P-1$. No two among these polynomials are equivalent. Furthermore, any polynomial obtained from $Q_0$ by replacing $x_1$ by $\eta^{k_1}x_1$ and every other variable $x_j$ by $\eta^{k_j}x_j$, where $k_j$ ranges over the same values as $k_1$, will be identical with one of the polynomials $Q_i$, say $Q_k$. The index $k$ is, namely, the smallest non-negative number which satisfies the congruence

$$y(m-q) \equiv (m-q)k_1 + k_2 + \cdots + k_s \pmod{P}.$$

We conclude, also, from the fact that $Q$ is primary in all its variables, that the same set of polynomials $Q_i$ will be obtained from $Q_0$ by means of every subgroup $g_j$ of substitutions which replace $x_j$ by $\eta^{k_j}x_j$, $k_j$ and $\eta$ having the same meanings as before, and leave all other variables unchanged.

This implies, first, that the product of the $P$ polynomials $Q_i$, which remains invariant for any substitution of our group, is an integral rational function $R(x_1^P, \cdots, x_s^P)$ in $x_j^P$, for every $x_j$. Moreover, since any of the substitutions of the subgroup $g_j$ simply permutes the $Q_i$'s, with respect to which $g_j$ is transitive, we conclude that $R(x_1^P, x_2^P, \cdots, x_j^P, x_{j+1}, \cdots, x_s)$, for every $j$ less than $s$, is irreducible. For, if such a polynomial were a product

$$R_1(x_1, \cdots, x_s)\cdot R_2(x_1, \cdots, x_s)$$

then

* According to Tchebycheff, given an integer $a$, greater than unity, there exists always a prime number between $a$ and $2a-1$. If, therefore, $P$ is the greatest prime contained in $m^2$, then $P>m^2/2$ and, consequently, $P>m$, considering that $m>1$.

$$R(x_1^P, \cdots, x_s^P) = R_1(x_1, \cdots, x_j, \overset{P}{x_{j+1}}, \cdots, x_s^P) \cdot R_2(x_1, \cdots, x_j, \overset{P}{x_{j+1}}, \cdots, x_s^P)$$

and none of the subgroups $g_k (k=j+1, \cdots, s)$ could, therefore, be transitive with respect to the polynomials $Q_i$.

The polynomial $R(x_1, \cdots, x_s)$ obtained from $R(x_1^P, \cdots, x_s^P)$ by replacing each $x_j^P$ by $x_j$, is, therefore, also irreducible. If $m_i$ is the degree of $Q_0$ in $x_i$, then $m_i P$ is the corresponding degree of $R(x_1^P, \cdots, x_s^P)$. Hence, the degree of $R(x_1, \cdots, x_s)$ in $x_i$ is $m_i$. Because $m_2 = m_3 = \cdots = m_s = m$ and both numbers $m-q$ and $(m-q)m-P$ are less than $m$, we see that $\max(m_1, \cdots, m_s) = M = m$. The irreducible polynomial $R(x_1, \cdots, x_s)$ becomes reducible when each $x_j$ is replaced by $x_j^P$. Hence the set $t_1 = P, \cdots, t_s = P$ is either a minimal set of $R(x_1, \cdots, x_s)$, or is obtained from a minimal set whose elements are, then, necessarily equal either to unity or to $P$. Suppose, fixing our ideas, that its elements $\tau_1 = \tau_2 = \cdots = \tau_j = P$ and $\tau_{j+1} = \cdots = \tau_s = 1$. This implies that $R(x_1^P, \cdots, x_j^P, x_{j+1}, \cdots, x_s)$ is reducible. We have seen, however, that a polynomial of this type is always irreducible unless $j = s$.

The polynomial $R(x_1, \cdots, x_s)$ has, therefore, a minimal set all the elements of which are equal to $P$, where $P$ is the largest prime less than $M$[2].

## III. Polynomials of two terms

16. We shall, for the sake of completeness, investigate the case in which the irreducible polynomial $Q$ has two terms. Changing subscripts, if necessary, we assume that

$$Q(x_1, \cdots, x_s) = a x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k} + b x_{k+1}^{\alpha_{k+1}} \cdots x_s^{\alpha_s}. \tag{37}$$

Let $t_1, \cdots, t_s$ be a set of positive integers for which $Q^{(t)}$ is reducible. We shall prove that in the identity

$$Q^{(t)} = Q_1 \cdots Q_n, \tag{38}$$

$Q_1$, and, on the basis of Lemma I*, each $Q_i$, contains only two terms.

Let $\lambda_i$ be the greatest common factor of the exponents of $x_i$ in $Q_1$. We conclude from (37) and (38) that $\alpha_i t_i$ must be divisible by $\lambda_i$. Hence if $\alpha_i t_i = \lambda_i \tau_i$ and if $P_k$ is the polynomial obtained from $Q_k$ by replacing each $x_j^{\lambda_j}$ by $x_j$, we find that

$$a x_1^{\tau_1} \cdots x_k^{\tau_k} + b x_{k+1}^{\tau_{k+1}} \cdots x_s^{\tau_s} = P_1 \cdots P_N, \tag{39}$$

where all the $P_i$'s are primary.

* In the proof of this lemma, as well as Lemma III, no restriction was made with regard to the number of terms in the irreducible polynomial $Q$.

If each $Q_i$, and, consequently, each $P_i$, consisted of more than two terms, then, as a consequence of Lemma III, $\tau_j \leqq M^2$. Because $M$ is unity, each $\tau_j$ is unity, which is an absurdity. Hence, each $Q_i$ contains only two terms.

Since $P_i$ is primary and consists of only two terms, its degree in $x_j$ is unity. Comparing the degrees in $x_j$ of both sides of the identity (39) we find that $\tau_j = N$, for every $j$.

Consider the infinite system of sets

$$(\gamma) \qquad t_{21}, \cdots, t_{2s}; \; t_{31}, \cdots, t_{3s}; \cdots$$

where $t_{kj} = k/d_{kj}$ and $d_{kj}$ is the greatest common factor of $\alpha_j$ and $k$.

We easily verify that, for any set of $(\gamma)$, the corresponding $Q^{(i)}$ is reducible.

Further, if, for a given set of $t$'s, $Q^{(i)}$ is reducible, then there exists in $(\gamma)$ one and only one set, say $t_{i1}, \cdots, t_{is}$, such that each $t_j$ is an integral multiple of $t_{ij}$ and such that if $t_j = \delta_j t_{ij}$, the irreducible factors of $Q^{(i)}$ are found by replacing in the irreducible factors of $Q(x_1^{t_{i1}}, \cdots, x_s^{t_{is}})$ each $x_j$ by $x^{\delta_j}$.

Finally, it can be shown that there exists no other system of sets satisfyin the two condition s just stated.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# FUNCTIONAL PROPERTIES OF THE SOLUTIONS OF DIFFERENTIAL SYSTEMS*

BY

WILLIAM M. WHYBURN

Numerous contributions have recently been made to the subject of functionals. The applications of work in this field are found in many branches of pure mathematics and mathematical physics. Bliss† developed functional properties of the solutions of a system of differential equations and made use of these properties in his work on ballistics. In the new field of wave mechanics a tool that is of extreme usefulness is a method of perturbations based entirely upon underlying functional properties of the solutions of differential equations. An investigation of stability in dynamical systems, especially stability of the perturbative type,‡ is essentially a study of the functional properties of a class of differential systems. Recently W. L. Hart§ has developed the notion of continuity for a functional of arguments that are summable functions of their variables.

The present paper considers a general class of systems of non-linear differential equations of the first order and shows that the solutions of these systems are *uniformly* continuous functionals of the non-linear functions and the initial values in a stronger sense than is ordinarily considered. Estimates are given for the magnitudes of the variations in the solutions in terms of the variations of the quantities upon which they depend. All of the variables and functions in the paper are supposed real.

Consider the system of ordinary differential equations

$$(1)\quad dy_i/dx = f_i(x, y_1, \cdots, y_n) + \sum_{j=1}^{j=n} A_{ij}(x, y_1, \cdots, y_n)y_j \quad (i = 1, \cdots, n),$$

where the following conditions are imposed:¶

A. For each fixed $x$ on $X$: $a \leqq x \leqq b$, the functions $f_i(x, y)$ and $A_{ij}(x, y)$ are continuous in $(y_1, \cdots, y_n)$ for all real values of $y_1, \cdots, y_n$.

B. For each fixed $(y_1, \cdots, y_n)$, the functions $A_{ij}(x, y)$ and $f_i(x, y)$ are summable in $x$ on $X$.

---

* Presented to the Society, June 21, 1929; received by the editors September 3, 1929.

† These Transactions, vol. 21 (1920), pp. 79–92, and p. 92ff.

‡ See Birkhoff, American Journal of Mathematics, vol. 49 (1927), pp. 1–38.

§ Annals of Mathematics, (2), vol. 24 (1922–23), pp. 23–38.

¶ $A_{ij}(x, y) = A_{ij}(x, y_1, \cdots, y_n)$, $f_i(x, y) = f_i(x, y_1, \cdots, y_n)$, etc.

C. There exist summable functions $M(x)$ and $L(x)$ on $X$ such that for all $x$ on $X$ and for all real $(y_1, \cdots, y_n)$, $(z_1, \cdots, z_n)$

$$\left\{\begin{array}{l} |\, f_i(x,y)\,| \\ |\, A_{ij}(x,y)\,| \end{array}\right\} \leqq M(x),$$

$$\left\{\begin{array}{l} |\, A_{ij}(x,y) - A_{ij}(x,z)\,| \\ |\, f_i(x,y) - f_i(x,z)\,| \end{array}\right\} \leqq L(x) \sum_{r=1}^{r=n} |\, y_r - z_r\,| \qquad (i,j = 1, \cdots, n).$$

The author has recently shown* that if $T$ is an arbitrarily given constant and if $\alpha_1, \cdots, \alpha_n, a_1, \cdots, a_n$ are constants such that

D. $|\alpha_i| < T$, $i = 1, \cdots, n$, and the points $x = a_i$, $i = 1, \cdots, n$, all lie on a single sub-interval $\Gamma$ of $X$ of length† $\delta$,

then there exists a unique solution of (1) which satisfies

$$y_i(a_i) = \alpha_i \qquad (i = 1, \cdots, n) \tag{2}$$

where by a solution of (1) we understand a set of absolutely continuous functions $y_1(x), \cdots, y_n(x)$ that satisfies (1) almost everywhere on $X$. This existence theorem is quite general and includes the usual existence and uniqueness theorems for systems of the type ordinarily considered, i.e., where $A_{ij} = 0$, $a_1 = a_2 = \cdots = a_n$, as special cases and in addition covers other systems, e.g., the linear system. We treat this more general system (1), (2) in such a way that the results for the ordinary systems occur as special cases of our results.

For a given $M(x)$, $L(x)$, and $T$ let $Q$ denote the class of functions $A_{ij}(x,y)$, $f_i(x, y)$, $\alpha_i$, $a_i$, $i, j = 1, \cdots, n$, that satisfy conditions A, B, C, D. Let $R$ denote the sub-class of $Q$ for which $a_1 = a_2 = \cdots = a_n = c$.

THEOREM I. *For given $M(x)$, $L(x)$, and $T$, there exists a constant $H$ such that if $A_{ij}, f_i, \alpha_i, a_i, i, j = 1, \cdots, n$, are any functions‡ of $Q$ and $y_1(x), \cdots, y_n(x)$ is the unique solution of* (1), (2), *then*

$$|\, y_i(x)\,| \leqq H \qquad (i = 1, \cdots, n,\ a \leqq x \leqq b).$$

An examination of the author's paper§ shows that the quantity $G$ there determined depends only on $T$, $L$, and $M$—the constants $r$ and $R$ having previously assigned constant values—and is therefore a constant for all $A_{ij}, f_i, \alpha_i, a_i$, in $Q$. When we choose a point $x = c$ of $\Gamma$ and examine the lemma of Theorem I of our previous paper,¶ we see that if $K$ of that lemma is chosen

* Annals of Mathematics, (2), vol. 30 (1928), pp. 31–38.

† $\delta$ is expressed in terms of $L$, $M$, and $T$ in the paper referred to.

‡ In case $A_{ij} \equiv 0$, $i, j = 1, \cdots, n$, $a_1 = a_2 = \cdots = a_n$, this theorem follows directly from an application of Theorem IV of the present paper.

§ Loc. cit., p. 35, Lemma 2 of Theorem II.

¶ Loc. cit., p. 32.

greater than $nG+\int_a^b M(t)dt$, we obtain a constant $H=K\exp[n\int_a^b M(t)dt]$ such that for any given set of quantities $A_{ij}, f_i, \alpha_i, a_i$, in $Q$, we have

$$|y_i^{(k)}(x)| \leqq H \qquad (i=1,\cdots,n,\ k=0,1,\cdots,\ a\leqq x\leqq b).$$

Since $y_i(x)$ is the limit of the sequence of functions $y_i^{(k)}(x)$, $k=0, 1, \cdots$, it follows that $|y_i(x)|\leqq H$, $i=1,\cdots,n$, $a\leqq x\leqq b$. Theorem I follows from the observation that $H$ is independent of the particular set of quantities $A_{ij}$, $f_i, \alpha_i, a_i$, of $Q$ that is used.

THEOREM II. *Let $u_{1j}(x),\cdots,u_{nj}(x)$, $j=1,\cdots,n$, be the solutions of*

$$du_i/dx = \sum_{j=1}^{j=n} F_{ij}(x)u_j,\ u_{ir}(c) = \begin{cases} 0, & i\neq r \\ 1, & i=r \end{cases} \quad (i=1,\cdots,n), \tag{3}$$

*where $x=c$ is a point of $X$, $F_{ij}(x)$ is summable on $X$, and let $J(x)$ be a summable function of $x$ on $X$ such that $|F_{ij}(x)|<J(x)$, $i,j=1,\cdots,n$. Let $W(x)$ be the determinant whose element in the $i$th row and $j$th column is $u_{ij}(x)$ and let $\beta_{ij}$ be the cofactor of $u_{ij}$ in $W$. There exists a quantity $N$ which is a constant for a given $J(x)$ and is such that if $K_{ij}=\beta_{ij}/W$, then $|K_{ij}|<N$ $(i,j=1,\cdots,n)$, $a\leqq x\leqq b$, and for all summable functions $F_{ij}(x)$ such that $|F_{ij}(x)|<J(x)$ on $X$.*

A well known theorem gives $W(x)=\exp[\int_a^x \sum_{i=1}^n F_{ii}(t)dt]$. Since $|F_{ii}(t)|<J(t)$ on $X$, we have $W(x)>\exp[-\int_a^b nJ(t)dt]>0$. Applying Theorem I to $u_{ij}(x)$, we get it less than a constant $S$, independent of $F_{ij}$, and hence each element of the determinant $\beta_{ij}$ is less than $S$. An application of Hadamard's theorem* yields $\beta_{ij}\leqq(nS^2)^{n/2}$, hence $|K_{ij}(x)|<N$, where

$$N = (S^2n)^{n/2}\exp\left[\int_a^b nJ(t)dt\right].$$

THEOREM III. *Let $F(x)$, $f_i(x)$, $g_i(x)$, $i=1,\cdots,n$, be finite-valued and measurable functions on the measurable point set $E$ and let $0\leqq F(x)\leqq\sum_{i=1}^n f_i(x)g_i(x)$ hold on $E$. There exist measurable functions $h_1(x),\cdots,h_n(x)$ such that $|h_i(x)|\leqq|g_i(x)|$ and $F(x)=\sum_{i=1}^n f_i(x)h_i(x)$ on $E$.*

Let $K_0$ be the subset of $E$ on which $F(x)=0$ and let $K_i$, $i=1,\cdots,n$, be the subset of $E$ on which $0, f_1g_1, f_1g_1+f_2g_2,\cdots, \sum_{j=1}^{i-1} f_jg_j<F(x)\leqq\sum_{j=1}^{i} f_jg_j$. The sets $K_i$ and $K_j$, $i\neq j$, have no common points, $g_i(x)f_i(x)$ is positive on $K_i$ for each $i=1,\cdots,n$, and $E=\sum_{j=0}^n K_j$. Let $h_1(x)\equiv 0$ on $K_0$, $h_1\equiv F(x)\cdot[f_1(x)]^{-1}$ on $K_1$, and $h_1\equiv g_1$ on $\sum_{j=2}^n K_j$. Let $h_i(x)$ be zero on $\sum_{j=0}^{i-1} K_j$, $h_i=[F(x)-\sum_{j=1}^{i-1} f_jg_j]/f_i(x)$ on $K_i$, and $h_i(x)\equiv g_i(x)$ on $\sum_{j=i+1}^n K_j$ $(i=2,\cdots,n)$.

* See Lovitt, *Linear Integral Equations*, New York, 1924, p. 31.

The construction shows immediately that $|h_i(x)| \leqq |g_i(x)|$ and $F(x) = \sum_{i=1}^{n} f_i h_i$ on $E$. Each of the point sets $K_i$ is measurable since $K_0$ is measurable and the measurability of $K_j$ follows from that of $K_0, \cdots, K_{j-1}$. For $K_j$ is the subset of the measurable set $E - \sum_{i=0}^{j-1} K_i$ on which the measurable function $\sum_{i=1}^{j} f_i g_i - F(x)$ is positive or zero. Since $f_i(x) \neq 0$ on $K_i$, we may apply a well known theorem* concerning products and quotients of measurable functions to get that $h_i$ is measurable on $K_i$. On $\sum_{j=i+1}^{n} K_j$, $h_i$ is equal to the measurable function $g_i(x)$ while $h_i \equiv 0$ on $\sum_{j=0}^{i-1} K_j$. Hence $h_i$ is measurable on $E$.

COROLLARY 1. *If $F(x) = |A(x)|$, where $A(x)$ is measurable on $E$, then $A(x) = \sum_{i=1}^{n} H_i(x) f_i(x)$, where $H_i(x) = +h_i(x)$ or $-h_i(x)$ according as $A(x) \geqq 0$ or $A(x) \leqq 0$. Hence $H_i(x)$ is measurable and $|H_i(x)| \leqq |g_i(x)|$ on $E$.*

COROLLARY 2. *If the functions $g_i(x)$ are summable on $E$, the functions $h_i(x)$ and $H_i(x)$ of Theorem* III *and Corollary* 1 *are summable on $E$.*†

THEOREM IV. *Let $f(x)$ be defined on any interval $X$ and let $M$ denote the subset of $X$ on which $df/dx$ exists (either finite or infinite with definite sign); then with the possible exception of a countable set of points $d|f(x)|/dx$ exists and $|d|f(x)|/dx| = |df(x)/dx|$.*

Let $x = p$ be any point of $M$ for which $f(p) \neq 0$. Since $f'(p)$ exists, $f(x)$ is continuous at $x = p$ and there exists a neighborhood of $x = p$ throughout which $f(x)f(p) > 0$. Throughout this neighborhood we have either $|f(x)| = f(x)$ or $|f(x)| = -f(x)$. In either case $d|f(p)|/dx$ exists and is numerically equal to $|f'(p)|$.

Let $x = p$ be a point of $M$ at which $f(x) = 0$. From the existence of $f'(p)$ we have $\lim_{h \to p^+} f(h)/(h-p) = f'(p)$ and hence $\lim_{h \to p^-} |f(h)|/(h-p) = -|f'(p)|$, $\lim_{h \to p^+} |f(h)|/(h-p) = |f'(p)|$. Hence both the left-hand and right-hand derivatives of $|f(x)|$ exist at every point of $M$. A theorem of G. C. Young‡ may now be applied to show that these two derivatives differ from each other at most at a countable set of points of $M$.

COROLLARY. *There does not exist a function that is defined on an interval and vanishes at more than a countable set of points at each of which its derivative exists (either finite or infinite with definite sign) and is different from zero.*§

* See Carathéodory, *Vorlesungen ueber reelle Funktionen*, Leipzig, 1928, p. 378.

† This follows since these functions are measurable and bounded numerically by the summable functions $|g_i(x)|$. See Carathéodory, loc. cit., p. 442.

‡ For a statement of this theorem, see Hobson, *Theory of Functions of a Real Variable*, Cambridge University Press, second edition, vol. I, pp. 368 and 369. The consequence of Young's theorem that applies directly in the above proof is stated at the bottom of p. 369.

§ This corollary follows from Theorem IV and its proof when we observe that the derivative of $|f(x)|$ does not exist at a point where $f(x)$ vanishes and has a derivative that is different from zero.

Equations (1), (2) define $y_1(x), \cdots, y_n(x)$ as functionals of $f_i(x, y)$, $A_{ij}(x, y)$, $\alpha_i$, $a_i$ $(i, j=1, \cdots, n)$, in $QX$ in the sense that for each set of functions in $Q$, the system (1), (2) determines unique absolutely continuous functions $y_1, \cdots, y_n$ on $X$ that satisfy (2) and satisfy (1) almost everywhere.

Following the definitions given by W. L. Hart,* we say that the functions $y_r(x)$ $(r=1, \cdots, n)$ are *continuous* functionals of $A_{ij}$, $f_i$, $\alpha_i$, $a_i$ *at* $B_{ij}$, $g_i$, $\beta_i$, $b_i$ of $Q$ if for every given positive number $\epsilon$, there exists a positive number $\delta_\epsilon$, such that if $A_{ij}$, $f_i$, $\alpha_i$, $a_i$ belong to $Q$ and $|a_i-b_i|<\delta_\epsilon$, $|\alpha_i-\beta_i|<\delta_\epsilon$, $\int_a |f_i(x, u)-g_i(x, u)|dx<\delta_\epsilon$, $\int_a^b |A_{ij}(x, u)-B_{ij}(x, u)|dx<\delta_\epsilon$, for all absolutely continuous functions $u_1(x), \cdots, u_n(x)$ that are bounded numerically on $X$ by $H$, the constant determined in Theorem I, $i, j=1, \cdots, n$, then $|y_r(x)-z_r(x)|<\epsilon$ for every $x$ on $X$, where $y_1(x), \cdots, y_n(x)$ and $z_1(x), \cdots, z_n(x)$, respectively, are the solutions of system (1), (2) corresponding to $A_{ij}$, $r_i$, $\alpha_i$, $a_i$, and $B_{ij}$, $g_i$, $\beta_i$, $b_i$. The continuity is said to be *uniform* in $Q$ if the number $\delta_\epsilon$ can be chosen independent of the functions $B_{ij}$, $g_i$, $\beta_i$, $b_i$, so long as they belong to $Q$.

THEOREM V. *The solution functions,* $y_1(x), \cdots, y_n(x)$, *of system* (1) *and*

$$y_i(c) = \alpha_i, \qquad a \leqq c \leqq b \qquad (i = 1, \cdots, n), \tag{4}$$

*are uniformly continuous functionals of* $A_{ij}, f_i, \alpha_i$, *in* $XR$.

Let a positive number $\epsilon$ be assigned and choose $\delta_\epsilon$ so that

$$\delta_\epsilon \leqq \epsilon/[nS\{1 + nN(1 + nH)\}],$$

where $N$ and $S$ are the constants whose existence was demonstrated in Theorem II for $F_{ij}(x) \equiv L(x)+M(x)+nHL(x)$ and $H$ is the constant whose existence is proved in Theorem I. Let $A_{ij}, f_i, \alpha_i$, and $B_{ij}, g_i, \beta_i$, be two sets of functions in $R$ and meeting the requirements of our definition of continuity with respect to $\delta_\epsilon$. Let $y_i(x)$ and $z_i(x)$, respectively, be the corresponding solutions of (1), (4), and let $u_i(x)=y_i(x)-z_i(x)$, $i=1, \cdots, n$. Almost everywhere on $X$ we have

$$\begin{aligned} u_i' &= f_i(x,y) - g_i(x,z) + \sum_{j=1}^{n}[A_{ij}(x,y)y_j - B_{ij}(x,z)z_j] \\ &= [f_i(x,y) - f_i(x,z)] + [f_i(x,z) - g_i(x,z)] + \sum_{j=1}^{n}\{[A_{ij}(x,y) \\ &\quad - B_{ij}(x,y)]y_j(x) + [B_{ij}(x,y) - B_{ij}(x,z)]y_j(x) + B_{ij}(x,z)[y_j - z_j]\}. \end{aligned} \tag{5}$$

Taking absolute values and making use of the inequalities of condition C and Theorem I, we have

* Loc. cit., p. 25.

$$| u_i' | \leqq [L(x) + M(x) + HnL(x)] \sum_{r=1}^{n} | u_r | + \phi_i(x) \qquad (i = 1, \cdots, n),$$

where

$$\phi_i(x) = | f_i(x,z) - g_i(x,z) | + H \sum_{j=1}^{n} | A_{ij}(x,y) - B_{ij}(x,y) |$$

and $y_i(x)$ and $z_i(x)$ are absolutely continuous functions of $x$ on $X$. Replacing $|u_i|$ by $U_i$ and making use of Theorem IV, we have

$$| U_i' | \leqq [L(x) + M(x) + nHL(x)] \sum_{r=1}^{n} U_r + \phi_i(x),$$

almost everywhere on $X$. Application of Theorem III and its corollaries yields

$$U_i' = C_i(x) \sum_{r=1}^{n} U_r + \theta_i(x), \quad U_i(c) = | \alpha_i - \beta_i | , \tag{6}$$

almost everywhere on $X$, where $C_i(x)$ and $\theta_i(x)$ are summable on $X$ and

$$| C_i(x) | \leqq [L + M + nHL], \quad | \theta_i(x) | \leqq \phi_i(x).$$

Solving (6) by a generalization of the method of variation of parameters yields

$$U_i(x) = \sum_{r=1}^{n} \left[ \sum_{j=1}^{n} \int_c^x \theta_j(t) K_{jr}(t)\, dt + | \alpha_r - \beta_r | \right] u_{ir}(x) \qquad (i = 1, \cdots, n),$$

where $u_{ij}(x)$ and $K_{ij}(x)$ are the quantities that occur in Theorem II. Applying Theorem II

$$\begin{aligned} U_i(x) &\leqq SnN \sum_{j=1}^{n} \int_c^x \theta_j(t)\, dt + S \sum_{r=1}^{n} | \alpha_r - \beta_r | \\ &< Sn^2N[1 + nH]\delta_\epsilon + nS\delta_\epsilon \leqq Sn[1 + nN(1 + nH)]\epsilon / \{Sn[1 + nN(1 + nH)]\} \\ &< \epsilon, \end{aligned}$$

which establishes the desired uniform continuity of $y_i(x)$, $i = 1, \cdots, n$.

COROLLARY. *The conclusions of Theorem* V *are valid for any solution of* (1), (4) *since* $T$ *may be chosen as large as one desires.*

In applications of functional properties of the solutions of (1), (4) it is of fundamental importance to be able to estimate the magnitude of the change in the solution functions, $y_i(x)$, that is produced by a change of known magnitude in $A_{ij}$, $f_i$, $\alpha_i$. The following theorem which is established im-

mediately by an examination of the proof of Theorem V gives such an estimate:*

THEOREM VI. *Let $A_{ij}, f_i, \alpha_i$ and $B_{ij}, g_i, \beta_i$ be two sets of functions in $R$ such that $|\alpha_i-\beta_i|<\delta$, $\int_a^b |A_{ij}(x,u)-B_{ij}(x,u)|dx<\delta$, $\int_a^b |f_i(x,u)-g_i(x,u)|dx<\delta$, for all absolutely continuous functions $u_i(x)$ on $X$, that are bounded numerically by $H$, $i, j=1, \cdots, n$, and let $y_1(x), \cdots, y_n(x)$ and $z_1(x), \cdots, z_n(x)$, respectively, denote the corresponding solutions of* (1), (4); *then*

$$|y_i(x)-z_i(x)|<Sn[1+nN(1+nH)]\delta \qquad (i=1, \cdots, n),$$

*on $X$, where $S, N$, and $H$ have the same significance as in the proof of Theorem* V.

A combination of Theorem V with the results of our former paper† yields the following more general theorem:

THEOREM VII. *The solution functions $y_1(x), \cdots, y_n(x)$ of system* (1), (2) *are uniformly continuous functionals of $A_{ij}, f_i, \alpha_i, a_i$, $i, j=1, \cdots, n$, in $QX$.*

By Theorem V and its corollary, $y_n(x)$ is a uniformly continuous functional on $X$ of $A_{ij}, f_i, \gamma_i$, where $\gamma_i$ is the value of $y_i(x)$ at a point $x=c$ of $\Gamma$. By our former theorem,‡ $\gamma_i$ is a continuous function of $(\alpha_1, \cdots, \alpha_n, a_1, \cdots, a_n)$ for $a_j$ on $\Gamma$ $(i, j=1, \cdots, n)$. It follows from this (by an argument that is entirely analogous to the well known proof that a continuous function of a continuous function is a continuous function) that $y_r(x)$ $(r=1, \cdots, n)$ is a continuous functional of $A_{ij}, f_i, \alpha_i, a_i$, in $QX$. It also follows from this argument and from the observation that our former theorem‡ yields uniformity of our type for $\gamma_r$ as a functional of $(\alpha_1, \cdots, \alpha_n, a_1, \cdots, a_n)$ that the continuity of the functionals $y_r(x)$ is uniform.

* It is to be noted that this estimate can be improved if some of the quantities $A_{ij}, f_i, \alpha_i$ are fixed. Obvious modifications of the proofs of the paper yield the results here.

† Whyburn, W. M., loc. cit., p. 37, Theorem III.

‡ Loc. cit., p. 37, Theorem III.

THE UNIVERSITY OF CALIFORNIA AT LOS ANGELES,
LOS ANGELES, CALIF.

# AN INVERSE PROBLEM OF THE CALCULUS OF VARIATIONS FOR MULTIPLE INTEGRALS*

BY

LINCOLN LA PAZ†

## 1. INTRODUCTION

Hirsch‡ has proved that if the equation of variation of the second-order partial differential equation

$$(1.1)\quad F(x,y,z,p,q,r,s,t) = 0,\quad z = z(x,y),\quad p = z_x,\ \cdots,\ t = z_{yy},$$

is self-adjoint, then the equation (1.1) has the form

$$(1.2)\quad M(rt - s^2) + Rr + 2Ss + Tt + N = 0,$$

where $M, R, S, T, N$ are functions of $x, y, z, p, q$ alone; and he has shown that in this case there exists a function $f$, itself of the second order, of the form§

$$(1.3)\quad f = g(x,y,z,p,q)s + c(x,y,z,p,q),$$

such that the equation (1.2) is the Lagrange partial differential equation of a problem of minimizing the double integral

$$(1.4)\quad J = \int\int f(x,y,z,p,q,s)dxdy.$$

A function $f$ thus related to a given partial differential equation with self-adjoint equation of variation will for brevity be referred to as *associated* with the given equation.

Kürschák,¶ utilizing the method of Hirsch, extended the latter's theorem to partial differential equations of the form

$$(1.5)\quad \begin{gathered} F(x_1,\cdots,x_n,z,p_1,\cdots,p_n,p_{11},p_{12},\cdots,p_{nn}) = 0, \\ p_i = \partial z/\partial x_i,\quad p_{ij} = \partial^2 z/\partial x_i \partial x_j. \end{gathered}$$

Neither Hirsch nor Kürschák attempted to find either the most general

* Presented to the Society, August 29, 1929, and February 22, 1930; received by the editors January 14, 1930.

† National Research Fellow in Mathematics, 1928–1929.

‡ A. Hirsch, Mathematische Annalen, vol. 49 (1897), p. 49. Loc. cit., §9.

§ The functions $g$ and $c$ in (1.3) must satisfy certain partial differential equations of the second order involving the coefficients of the equation (1.2).

¶ J. Kürschák, Mathematische Annalen, vol. 60 (1905), p. 157.

function of the *second* order associated with a given equation or the most general *first*-order function associated with a given equation linear (necessarily) in the second partial derivatives. Koenigsberger,* without the use of the self-adjoint theory, obtained a formula for the integrand function $f$ of a problem

$$I = \iint f(x,y,z,p,q)dxdy, \tag{1.6}$$

having a given partial differential equation (1.1), linear in $r$, $s$, $t$, as its Lagrange equation; but his formula does not give the most general such function.

The determination of the most general second-order function associated with a given equation (1.5) is rendered irregular by the fact that an initial restriction must be placed on the form of such a function in order to insure the degeneration of the corresponding Lagrange partial differential equation from the fourth to the second order. This case is not treated in the present paper which is devoted to a determination of the most general first-order integrand function $f(x_1, \cdots, x_n, z, p_1, \cdots, p_n)$ associated with a partial differential equation (1.5) linear in the $p_{ij}$. The argument of §3 leading to this determination is based on the self-adjoint theory and is in part similar to that employed by Davis† in connection with an analogous question for simple integrals

$$I = \int_{x_1}^{x_2} f(x,y,z,y',z')dx, \tag{1.7}$$

having as their Euler equations a given pair of ordinary differential equations with self-adjoint equations of variation.

In the last two sections application of the results obtained is made to secure a generalization of Dirichlet's principle and to treat the problem of plane geometries in which straight lines are shortest.

## 2. Properties of the Lagrange equation

For a problem of minimizing the multiple integral

$$I = \int_{(n)} f(x_1, \cdots, x_n, z, p_1, \cdots, p_n)dx_1 \cdots dx_n, \tag{2.1}$$

the Lagrange partial differential equation has the form

$$f_{p_a z_a} + f_{p_a z} p_a + f_{p_a p_b} p_{ab} - f_z = 0. \tag{2.2}$$

* L. Koenigsberger, Sitzungsberichte der Akademie der Wissenschaften, Berlin, 1905, p. 250.

† D. R. Davis, these Transactions, vol. 30 (1928), p. 717.

Here, as elsewhere in this paper, the repeated subscripts $a$ and $b$ are umbral indices with the ranges 1 to $n$.

Consider a one-parameter family of integral surfaces, $z=z(x_1, \cdots, x_n, k)$, of the partial differential equation (2.2). The quantity $u$ defined by $u=\partial z(x_1, \cdots, x_n, k)/\partial k$ is called the variation of the family on the surface corresponding to the value $k$ of the parameter. On this surface $u$ satisfies the equation of variation of (2.2) which can be put in the form

$$A(u) = (\partial f_{p_a z}/\partial x_a - f_{zz})u + (\partial f_{p_a p_b}/\partial x_a)u_b + f_{p_a p_b}u_{ab} = 0, \tag{2.3}$$

where the differentiations with respect to the $x_i$ are partial when $z$ and its derivatives are considered as functions of the $x$'s, and where $u_b=\partial u/\partial x_b$, $u_{ab}=\partial^2 u/\partial x_a \partial x_b$. From (2.3) it can be shown that $A(u)$ and its adjoint $A'(u)$* are identical.

The Lagrange partial differential equation of a problem (2.1) is therefore an equation of the form

$$A_{ab}p_{ab} + B = 0, \tag{2.4}$$

with an equation of variation which is self-adjoint on every hypersurface $z=z(x_1, \cdots, x_n)$. In (2.4), $A_{ij}$, $B$ are functions of the variables $x_i$, $z$, $p_i$ alone and $A_{ij}=A_{ji}$.

A first necessary condition for the equation of variation of a partial differential equation (1.5) to be self-adjoint is that $F$ satisfy the relations†

$$2\partial F/\partial p_i - \sum_{k=1}^{n}(\partial/\partial x_k)(\partial F/\partial p_{ik})(1+\delta_k^i) = 0 \quad (i = 1, \cdots, n), \tag{K}$$

where $\delta_k^i$ is the Kronecker delta. When $F$ has the special form (2.4) these conditions become

$$(A_{ib})_{x_b} + (A_{ib})_z p_b + \{(A_{ib})_{p_a} - (A_{ab})_{p_i}\}p_{ab} - B_{p_i} = 0 \ (i = 1, \cdots, n). \tag{2.5}$$

Hence, since these relations are required to hold on every hypersurface $z=z(x_1, \cdots, x_n)$, we have

* The notation is that of Frobenius, Crelle's Journal, vol. 85, p. 207. The actual computation of $A'(u)$ can be avoided by the following artifice similar to one used by Hadamard (*Leçons sur le Calcul des Variations*, p. 319) in connection with ordinary linear differential expressions of the second order. Introduce a quantity $w$ defined by the equation

$$2w=f_{zz}u^2+2f_{zp_a}uu_a+f_{p_ap_b}u_au_b;$$

then it can be shown that

$$uA(v)-vA(u)=\partial(u\partial w/\partial v_a' - v\partial w/\partial u_a')/\partial x_a.$$

† Kürschák, loc. cit., p. 160. The writer has found that the conditions (K) are also sufficient conditions for self-adjointness.

THEOREM I. *An equation* (1.5) *which as it stands is to be the Lagrange partial differential equation of a problem* (2.1) *must necessarily be of the form* (2.4) *with coefficients* $A_{ij}$, $B$ *satisfying the conditions*

$$(2.6)\qquad \begin{aligned} &A_{ij} = A_{ji}, \quad (A_{ij})_{p_k} = (A_{kj})_{p_i} && (i,j,k = 1, \cdots, n), \quad k \neq i, \\ &(A_{ib})_{z_b} + (A_{ib})_z p_b - B_{p_i} = 0 && (i = 1, \cdots, n), \end{aligned}$$

*identically in the variables* $x_i$, $z$, $p_i$.

## 3. DETERMINATION OF THE MOST GENERAL FIRST-ORDER INTEGRAND FUNCTION

Suppose that an equation (2.4) with coefficients satisfying the conditions (2.6) is given. In view of the relations in the first row of (2.6) we can define a function $g(x_1, \cdots, x_n, z, p_1, \cdots, p_n)$ by means of the system of equations

$$(3.1)\qquad g_{p_i p_j} = A_{ij} \qquad (i,j = 1,2, \cdots, n),$$

and the function $g$ thus defined can be expressed as a line integral

$$(3.2)\qquad g = \int_0^p H_a dp_a,$$

where

$$(3.3)\qquad H_i = \int_0^p A_{ib} dp_b \qquad (i = 1,2, \cdots, n).$$

In (3.2) and (3.3) the symbols 0 and $p$, used as limits, represent $n$-partite variables. It is clear that if $G$ is a particular solution of (3.1) then the most general solution of this system is of the form

$$(3.4)\qquad \begin{aligned} g = G(x_1, \cdots, x_n, z, p_1, \cdots, p_n) \\ + C(x_1, \cdots, x_n, z) + D_a(x_1, \cdots, x_n, z)p_a. \end{aligned}$$

If, now, the given equation (2.4) is to be the Lagrange equation of a problem (2.1) formed with the function $g$ in (3.4) as the integrand of the multiple integral, then the following relation must hold identically in the variables $x_i$, $z$, $p_i$, $p_{ik}$:

$$(3.5)\qquad \partial g_{p_a}/\partial x_a - g_z = A_{ab}p_{ab} + B.$$

Reduction of (3.5) by means of (3.1) shows that the relation below must hold identically in the variables $x_i$, $z$, $p_i$:

$$(3.6)\qquad C_z - D_{az_a} = G_{p_a z_a} + G_{p_a z}p_a - G_z - B.$$

If there are to exist functions $C$, $D_i$ $(i = 1, \cdots, n)$ of $x_1, \cdots, x_n$ and $z$ alone

satisfying (3.6), then the right member of this equation must be independent of $p_i$ $(i=1, \cdots, n)$. That this condition is actually fulfilled is seen by differentiating the right side of (3.6) with respect to $p_i$ and referring to the set of relations in the second row of (2.6).

Solutions of (3.6) for $C$, $D_i$ as functions of $x_1, \cdots, x_n, z$ alone therefore always exist, and if the set $(C, D_i)$ gives a particular solution of this equation then the most general solution is given by the set $(C+c, D_i+d_i)$ where $(c, d_i)$ are functions of $x_1, \cdots, x_n, z$ satisfying the equation

$$c_z - d_{az_a} = 0, \tag{3.7}$$

but otherwise arbitrary.

The relation (3.7) is a necessary and sufficient condition for the Lagrange partial differential equation of a problem of minimizing the multiple integral

$$\int_{(n)} (c + d_a p_a) dx_1 \cdots dx_n \tag{3.8}$$

to be identically satisfied. From this it follows, in view of a theorem of Koenigsberger* on the conditions under which the Lagrange equation can be identically satisfied, that the quantity $c+d_ap_a$ is representable as a sum of partial† derivatives with respect to the $x_i$ of arbitrary functions $w_i$ of $x_1, \cdots, x_n, z$ alone

$$c + d_a p_a = \partial w_1/\partial x_1 + \cdots + \partial w_n/\partial x_n. \tag{3.9}$$

This completes the proof of

THEOREM II. *If an equation $F(x_1, \cdots, x_n, z, p_1, \cdots, p_n, p_{11}, \cdots, p_{nn})=0$ is linear in the $p_{ij}$ and has an equation of variation which is self-adjoint on every hypersurface $z=z(x_1, \cdots, x_n)$ then there is always a multiple integral of the form*

$$I = \int_{(n)} f(x_1, \cdots, x_n, z, p_1, \cdots, p_n) dx_1 \cdots dx_n$$

*having $F=0$ as its Lagrange partial differential equation. The most general such integral has an integrand function of the form*

$$f = G + C + D_a p_a + \partial w_1/\partial x_1 + \cdots + \partial w_n/\partial x_n, \tag{3.10}$$

*where $G(x_1, \cdots, x_n, z, p_1, \cdots, p_n)$ is a particular solution of* (3.1), *where the set $\{C(x_1, \cdots, x_n, z), D_i(x_1, \cdots, x_n, z)\}$ is a particular solution of* (3.6) *and where the $w_i$ occurring in the partial derivatives with respect to $x_i$ are arbitrary functions of $x_1, \cdots, x_n, z$ alone.*

* Koenigsberger, Mathematische Annalen, vol. 62, p. 123.

† When $z$ is regarded as a function of the $x$'s.

As an immediate corollary we have a theorem analogous to one stated by Hirsch for simple integrals:*

THEOREM III. *If the Lagrange partial differential equation of a problem of minimizing a multiple integral with second-order integrand function*

$$\int_{(n)} f_2(x_1, \cdots, x_n, z, p_1, \cdots, p_n, p_{11}, \cdots, p_{nn})dx_1 \cdots dx_n$$

*degenerates into a partial differential equation, $L=0$, of the second order and linear in the $p_{ij}$, then by means of quadratures a first-order integrand function $f(x_1, \cdots, x_n, z, p_1, \cdots, p_n)$ can be determined such that a problem of minimizing the integral*

$$\int_{(n)} f_1(x_1, \cdots, x_n, z, p_1, \cdots, p_n)dx_1 \cdots dx_n$$

*leads to the same partial differential equation, $L=0$.*

## 4. A GENERALIZATION OF DIRICHLET'S PRINCIPLE

The partial differential equation

$$F = \partial^2 z/\partial x_1^2 + \partial^2 z/\partial x_2^2 + \cdots + \partial^2 z/\partial x_n^2 = 0 \tag{4.1}$$

satisfies the conditions of Theorem II of the last section. The system (3.1) is in this case

$$g_{p_i p_j} = \delta_j^i. \tag{4.2}$$

Hence we may take for the $G$ function in (3.10)

$$G = \tfrac{1}{2}(p_1^2 + \cdots + p_n^2). \tag{4.3}$$

By means of (4.1) the relation (3.6) reduces to

$$C_\alpha - D_{\alpha x_\alpha} = 0. \tag{4.4}$$

Hence the sum $C + D_\alpha p_\alpha$ in (3.10) can itself be represented in the form (3.9). We conclude that the most general $n$-fold integral which has (4.1) as it stands as its Lagrange equation has an integrand function of the form

$$f = \tfrac{1}{2}(p_1^2 + \cdots + p_n^2 + \partial W_1/\partial x_1 + \cdots + \partial W_n/\partial x_n), \tag{4.5}$$

where the functions $W_i$ are arbitrary functions of $x_1, \cdots, x_n$ and $z$.

The expression in parentheses can be written in the form

$$\mathfrak{F}(x_1, \cdots, x_n, z, p_1, \cdots, p_n) = \sum_{i=1}^{n}(p_i + W_{iz}/2)^2 + (W_{ix_i} - W_{iz}^2/4). \tag{4.6}$$

* Hirsch, loc. cit., pp. 52–53.

A generalization of Dirichlet's principle would now assert that if the function $\mathfrak{F}$ is never negative in a region $S$ which consists of all sets of values $(x_1, \cdots, x_n, z, p_1, \cdots, p_n)$ where $(x_1, \cdots, x_n, z)$ is a point in a certain domain $D$ of $(x_1, \cdots, x_n, z)$-space and where $p_1, \cdots, p_n$ have arbitrary finite values, then there actually exists a hypersurface $z=z(x_1, \cdots, x_n)$ minimizing the $n$-fold integral

$$\mathfrak{D} = \int_{(n)} \mathfrak{F} dx_1 \cdots dx_n, \tag{4.7}$$

at least in $D$.

For $n=2$ the partial differential equation (4.1) becomes the Laplacian, and there is thus established an interesting relation between the class of analytic functions of a complex variable and the class of real functions $z(x, y)$ which furnish solutions of a problem of minimizing double integrals of the form

$$\iint (p^2 + q^2 + \partial W_1/\partial x + \partial W_2/\partial y) dx dy. \tag{4.8}$$

The author hopes to investigate in a later paper the corresponding relationships for problems of the calculus of variations having as their Lagrange equation the generalized Laplace equation by means of which Hedrick and others have classified functions in the plane.*

## 5. Geometries in which straight lines are the curves of shortest length

As a further application of the results obtained in Theorems I and II, we consider the problem, first treated by Hamel in his Dissertation, of determining the most general geometry in the plane for which straight lines are the curves of shortest length joining two given points.† It will be shown that Hamel did not obtain the most general such geometry.

Denote by $R$ the region which consists of all sets of values $(x, z, z')$ where $(x, z)$ is a point in the $xz$-plane and $z'$ has an arbitrary finite value. Denote by $H$ the class of *real* functions $h(x, z, z')$ which are positive and single-valued in $R$ and which together with their partial derivatives of the first three orders are continuous in this region.‡

Then formulated as an inverse problem of the calculus of variations Hamel's problem requires the determination of the most general function,

* Hedrick, Ingold and Westfall, Journal de Mathématiques, (9), vol. 2 (1923), p. 338, §10.

† Hamel, Mathematische Annalen, vol. 57 (1903), pp. 235–244.

‡ Functions of class $H$ satisfy all of the conditions imposed by the geometric axioms chosen by Hamel and the definition of length adopted by him. In Bolza's terminology the functions $h(x, z, z')$ are of class $C'''$ in $R$. See *Vorlesungen über Variationsrechnung*, p. 13.

$g(x, z, z')$, in the class $H$ such that the straight line segment $E_{12}$ joining the arbitrary but fixed points $1:(x_1, z_1)$ and $2:(x_2, z_2)$, $x_2 \neq x_1$, renders the length integral

$$L = \int_{x_1}^{x_2} g(x,z,z')dx \tag{5.1}$$

*a proper,** *absolute* minimum with respect to all other admissible† arcs, $z=z(x)$, joining these two points.

If the integral (5.1) is to have even a strong *relative* minimum on the straight line segment $E_{12}$, then $E_{12}$ must satisfy the necessary conditions of Euler, Weierstrass and Legendre.‡ Since the minimizing arcs are now prescribed these conditions must be regarded as necessary conditions on the integrand function $g(x, z, z')$ in (5.1). Thus the first or Euler condition is seen to require that the function $g(x, z, z')$ be so specialized that the two-parameter family of straight lines in the $xz$-plane is the extremal system of (5.1). Our Theorem I, written for $n=1$, shows that the most general equation $F(x, z, z', z'')=0$, *linear* in $z''$, with an equation of variation self-adjoint on every curve $z=z(x)$, which is equivalent to an equation of the form

$$z'' = G(x,z,z'), \tag{5.2}$$

is of the form

$$F = Pz'' - PG = 0, \tag{5.3}$$

where the multiplier $P(x, z, z')$ of class $C'$ is $\not\equiv 0$ and satisfies the relation

$$P_z + P_{z'}z' + P_{z'}G + PG_{z'} = 0. \tag{5.4}$$

The straight lines in question satisfy (5.2) with $G=0$. The most general solution of (5.4), when $G=0$, is $P=P(z', z-xz')$, where $P$, except for the restrictions just mentioned, is an arbitrary function of its arguments. Hence our problem requires the determination of the most general integrand function $g(x, z, z')$ such that the Euler equation of (5.1) is

$$P(z', z - xz')z'' = 0. \tag{5.5}$$

From Theorem II with $n=1$ the most general such function is seen to be

$$g = C(x,z,z') + D(x,z) + E(x,z)z' + du(x,z)/dx, \tag{5.6}$$

* Compare Hilbert, Bulletin of the American Mathematical Society, vol. 8 (1902), p. 450.

† For the meaning of this term and the significance of the region $R$ introduced above, see Bliss, *Calculus of Variations*, 1925, pp. 128–129.

‡ Bliss, loc. cit., pp. 130–132. The field condition is automatically satisfied since the limits of integration are fixed and the one-parameter family of straight lines through a point has no second envelope.

where $C$ is a particular solution of

$$C_{z'z'} = P(z', z - xz'), \tag{5.7}$$

where the set $(D, E)$ is a particular solution of

$$D_z - E_x = C_{z'x} + C_{z'z}z' - C_z, \tag{5.8}$$

and where $u$ is an arbitrary function of $x$ and $z$. The value for $C$ determined from (5.7) and the relation (5.4) written for $G=0$, enable us to show that the right side of (5.8) is identically zero. Therefore $D + z'E$ is itself the derivative of an arbitrary function of $x$ and $z$. Hence

$$g(x,z,z') = \int_{z_0'}^{z'} \int_{z_0'}^{z'} P(z', z - xz')dz'dz' + dw/dx, \tag{5.9}$$

where it is understood that the arbitrary elements $z_0'$, $P(z', z-xz')$ and $w(x, z)$ are to be so chosen that $g$ is a member of the class $H$.*

For functions $P(z', z-xz')$ such that the vanishing of $P$ implies the relation $z''=0$, the argument now proceeds in the original region $R$; for functions not of this character the argument is made in the largest subregion $R_1$ of $R$ in which $P \neq 0$ and in which the sets $(x, z, z')$ with $z_1' < z' < z_2'$ are admissible if the sets $(x, z, z_1')$ and $(x, z, z_2')$ possess this property. The conditions of Weierstrass and Legendre, since they must hold for arbitrary line segments $E_{12}$, require respectively

$$E(x,z,z',Z') \geqq 0, \tag{5.10}$$

at every admissible element† $(x, z, z')$, for every admissible element $(x, z, Z') \neq (x, z, z')$, and

$$g_{z'z'} \geqq 0, \tag{5.11}$$

at every admissible element $(x, z, z')$. In view of the nature of the regions $R$ and $R_1$, it is clear that we may in every case employ the relation‡

$$E(x,z,z',Z') = \tfrac{1}{2}(Z' - z')^2 g_{z'z'}(x,z,z' + \theta(Z' - z')),\ 0 < \theta < 1, \tag{5.12}$$

---

* Such an agreement is necessary since the integral (5.1) formed with an integrand $g(x, z, z')$ determined out of (5.9) may be minimized by the straight lines in the $xz$-plane without $g$ belonging to $H$. This fact is shown by the following example which exhibits a function $g$ not a member of $H$, for which (5.1) is minimized by the lines in question: Choose $z_0'=0$, $P=1$, $w=-xz$; then from (5.9) $g=\frac{1}{2}[z'(z'-2x)-2z]$. Hence in every point on a parabola of the family $z-a=x^2$, $a>0$, $g$ has the value $-z$ and is therefore $<0$. Hamel, who obtained for $g(x, z, z')$ the same formula (5.9) by the method of Darboux, imposes no condition on $z_0'$, $P$, or $w$ which excludes the geometry corresponding to the above example, in which arcs of the parabolas specified have a *negative* length.

† Compare second footnote p. 516.

‡ Bliss, loc. cit., p. 139, formula (13).

to show that the condition (5.11) implies the condition (5.10). Consequently if $g$ is so specialized that (5.11) is satisfied it will follow that the $E$-function is $\geqq 0$ in the field* $F$ consisting of the straight lines parallel to $E_{12}$ and this is sufficient to insure that $\Delta L = L(C_{12}) - L(E_{12}) > 0$ on every admissible arc $C_{12} \neq E_{12}$ unless the $E$-function can vanish otherwise than ordinarily in $F$.† Hence, sufficient conditions that every straight line segment $E_{12}$ in the plane with an equation of the form $z = z(x)$ shall furnish a strong, proper, relative minimum with respect to all admissible arcs $C_{12}$ joining the points 1 and 2 are that $g$ have the form (5.9) and that in the region under consideration

$$g_{z'z'} = P(z', z - xz') \geqq 0, \tag{5.13}$$

where functions $P$ for which the equality sign holds in (5.13) are to be excluded *if and only if* the $E$-function computed for the corresponding integrand function (5.9) can vanish for $Z' \neq z'$. Since every pair of points 1 and 2, $x_1 \neq x_2$, in the plane can be joined by one and only one straight line segment $E_{12}$, the above conditions are also sufficient‡ conditions for (5.1) to have a proper, absolute minimum on every $E_{12}$ of the type described. We thus reach

THEOREM IV. *If a geometry in which length is defined by the integral* (5.1) *has an arc element of the form*

$$dL = \left\{ \int_{z_0'}^{z'} \int_{z_0'}^{z'} P(z', z - xz')dz'dz' + dw(x,z)/dx \right\} dx, \tag{5.14}$$

*where the arbitrary elements $z_0'$, $P(z', z - xz')$, $w(x, z)$ are such that the quantity in brackets belongs to the class $H$ and where $P$ is never negative, with the understanding that functions $P$ which vanish are to be excluded if and only if the corresponding $E$-function vanishes extraordinarily; then the geometry is one for which straight lines are the curves of shortest length.*

A simple example now shows that Hamel did not obtain the most general geometry in which straight lines are shortest. Choose $z_0' = 0$, $P = 12z'^2$, and $w(x, z) = x$. Then from (5.9) $g = 1 + z'^4$. This function belongs to the class $H$. Furthermore it is readily verified that the corresponding $E$-function

$$E(x,z,z',Z') = Z'^4 - z'^4 - 4(Z' - z')z'^3 \tag{5.15}$$

is always $\geqq 0$ and vanishes for $Z' = z'$ and for no other *real* value of $Z'$. Hence the geometry for which the element of arc is $dL = (1 + z'^4)dx$ is one of

* Compare Bliss, loc. cit., pp. 27–30.

† Kneser, *Lehrbuch der Variationsrechnung*, 1900, §22. Bolza, *Vorlesungen über Variationsrechnung*, pp. 120–121. Bliss, loc. cit., p. 133, p. 153.

‡ Sets of conditions at once necessary and sufficient are exceptional in the calculus of variations.

the geometries satisfying the assumptions of Hamel, since for it the function $g$ belongs to the class $H$ and since for this geometry every straight line segment $E_{12}$ in the plane with an equation $z=z(x)$ is shorter than every other admissible arc joining the points 1 and 2. However, the geometry here described would have to be excluded if we argued incorrectly with Hamel* that a necessary condition for a minimum is that the $E$-function be always $>0$ and therefore that the condition $g_{z'z'}>0$ must always be satisfied. For the special geometry just described $P=g_{z'z'}=12z'^2$ and vanishes for $z'=0$.

UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

* Hamel, loc. cit., p. 243. Hamel's $E$ is the negative of our $E$-function. He writes "$\cdots$ $E$ muss stets ein negatives Vorzeichen besitzen."

# POSTULATES FOR THE INVERSE OPERATIONS IN A GROUP*

BY
MORGAN WARD

1. **Introduction.** Suppose that we are given a collection of marks which we shall call "$\mathfrak{S}$-symbols," and a rule which assigns to any two $\mathfrak{S}$-symbols $x$ and $y$ a unique third $\mathfrak{S}$-symbol $z$. We may then regard $z$ as a one-valued "function" of $x$ and $y$ defined over the collection $\mathfrak{S}$ and write

$$z = F(x, y). \tag{1}$$

It may happen that the function $F$ is of such a character that when the $\mathfrak{S}$-symbols $z$, $x$ are given in (1), an $\mathfrak{S}$-symbol $y$ is uniquely determined, and when the $\mathfrak{S}$-symbols $y$, $z$ are given in (1), an $\mathfrak{S}$-symbol $x$ is uniquely determined. In this case we may associate with the function $F(x, y)$ two other one-valued functions $y=G(z, x)$, $x=H(y, z)$ defined over the collection $\mathfrak{S}$. These functions are called the first and second *inverses* of the function $F$. The primary object of this paper is to state the restrictions which must be imposed upon $F$ in order that one of its inverses may define with $\mathfrak{S}$ an abstract group.

It is convenient, in developing the properties of the system $\{\mathfrak{S}; \circ\}$ consisting of the $\mathfrak{S}$-symbols, $F$, and one or more postulates, to replace (1) by the notation†

$$z = x \circ y.$$

In any interpretation of the $\mathfrak{S}$, we may look upon $\circ$ as an *operation* which we perform upon $x$ and $y$ to obtain $z$. We shall continue to call $\circ$ an operation even when no interpretation of the $\mathfrak{S}$-symbols is in mind, and we shall similarly replace $y=G(z, x)$ and $x=H(y, z)$ by

$$y = z \Delta x \quad \text{and} \quad x = y \square z.$$

For example, suppose that the $\mathfrak{S}$-symbols stand for the rational integers, and that $F(x, y)=x-y$. Then $z=x-y$; $y=-z+x$; $x=y+z$, so that $\circ$ is subtraction, $\square$, addition, and $\Delta$, the negative of subtraction. Our problem in this instance would be first of all to frame a definition of "subtraction," and then to define "addition" in terms of "subtraction."

* Presented to the Society, April 5, 1930; received by the editors of the Bulletin in January, 1929, and transferred to the Transactions.

† Read "$z$ equals $x$ dot $y$."

2. **Postulates for the operation** $\circ$. Consider first the system $\{\mathfrak{S}; \circ\}$ consisting of (i) a collection $\mathfrak{S}$ of two or more distinct elements $a, b, c, \cdots$, (ii) a function $F(x, y) = x \circ y$ defined over $\mathfrak{S}$, and (iii) the following four postulates:

POSTULATE 1. *If $a, b$ are any elements of $\mathfrak{S}$, then $a \circ b$ is an element of $\mathfrak{S}$ uniquely determined by $a$ and $b$.*

POSTULATE 2. *If $a, b$ are any elements of $\mathfrak{S}$, then $a \circ a = b \circ b$.*

POSTULATE 3. *If $a, b, c$ are any elements of $\mathfrak{S}$, and 1 is the element of definition 1 below, $(a \circ b) \circ c = a \circ (c \circ (1 \circ b))$.*

POSTULATE 4. *If $a, b$ are any elements of $\mathfrak{S}$, and 1 is the element of Definition 1 below, and if $1 \circ a = 1 \circ b$, then $a = b$.*

THEOREM 1. *There exists a unique element $i$ of $\mathfrak{S}$ such that $i \circ i = i$.*

By Postulate 1, $a \circ a = i$ is an element of $\mathfrak{S}$, and by Postulate 2, $i \circ i = a \circ a = i$. Moreover, if $j$ were any second element such that $j \circ j = j$, then by Postulate 2,

$$j = j \circ j = i \circ i = i.$$

DEFINITION 1. *The element $i$ of Theorem 1 is called the "identity" of $\mathfrak{S}$ and denoted by 1, so that Theorem 1 states that for any element $a$ of $\mathfrak{S}$*

$$a \circ a = 1 \circ 1 = 1.$$

Postulates 1, 3 and 4 are true in any Abelian group, or any Abelian semi-group containing an identity. Postulate 4 is in fact a weakened form of Dickson's third postulate for a semi-group.* Postulate 2 is far more drastic, and serves to give the system its peculiar character.

3. **Consistency and independence of postulates.** The consistency and independence of the four postulates given in §2 is proved by the following table, which gives examples of systems in which Postulates 1–4 are all true, Postulate 1 false and Postulates 2, 3, 4 true and so on.

It should be noted that in order to prove that Postulate 1 is independent, we must change the statement of the remaining postulates slightly. Thus Postulate 2 should read *If $a, b$ are any two elements of $\mathfrak{S}$, and if $a \circ a$, $b \circ b$ are both in $\mathfrak{S}$, then $a \circ a = b \circ b$.* The similar emendations of Postulate 3 and Postulate 4 are left to the reader.

---

* L. E. Dickson, *On semi-groups and the general isomorphism between infinite groups*, these Transactions, vol. 6 (1905), p. 205. Also see Theorem 5, §4, Theorem 10, §6.

TABLE I

| | | $\mathfrak{S}$ | $x \circ y$ |
|---|---|---|---|
| | Consistency | | |
| 1. | All four true | Rational integers | $x-y$ |
| 2. | All four true | Rationals, 0 excluded | $x \div y$ |
| | Independence | | |
| 3. | Postulate 1 false | Rational integers, 2 excluded | $x-y$ |
| 4. | Postulate 2 false | Rational integers | $x+y$ |
| 5. | Postulate 3 false | Rational integers | $y-x$ |
| 6. | Postulate 4 false | Rationals, 0 excluded | $\lvert x \rvert \div \lvert y \rvert$ |

4. **Deductions from postulates.** The five theorems which follow give important properties of the system $\{\mathfrak{S}; \circ\}$ defined in §2, and lead up to the fundamental theorem of the next section. The proofs of the theorems are given in some detail in order to bring out clearly the implications of the various postulates.

THEOREM 2. *If $a$ is any element of $\mathfrak{S}$, then $a \circ 1 = a$.*

We have $1 \circ a = (1 \circ 1) \circ a = 1 \circ (a \circ (1 \circ 1)) = 1 \circ (a \circ 1)$ by Theorem 1 and Postulate 3. Therefore, by Postulate 4, $a = a \circ 1$.

THEOREM 3. *If $a$ is any element of $\mathfrak{S}$, then $1 \circ (1 \circ a) = a$.*

We have $1 \circ a = (1 \circ a) \circ 1 = 1 \circ (1 \circ (1 \circ a))$, by Theorem 2 and Postulate 3. Therefore, by Postulate 4, $a = 1 \circ (1 \circ a)$.

THEOREM 4. *If $a$, $b$, $c$ are any three elements of $\mathfrak{S}$, then $a \circ (b \circ c) = (a \circ (1 \circ c)) \circ b$.*

We have $a \circ (b \circ c) = a \circ [b \circ (1 \circ (1 \circ c))] = (a \circ (1 \circ c)) \circ b$, by Theorem 3 and Postulate 3.

THEOREM 4.1. *If $a$, $b$ are any elements of $\mathfrak{S}$, then $1 \circ (a \circ b) = b \circ a$.*

We have $1 \circ (a \circ b) = (1 \circ (1 \circ b)) \circ a = b \circ a$ by Theorem 4 and Theorem 3.

THEOREM 5. *If $a$, $b$, $c$ are any elements of $\mathfrak{S}$ and if* (i) $b \circ a = c \circ a$, *then $b = c$; and if* (ii) $a \circ b = a \circ c$, *then $b = c$.*

(i) is clear from Postulate 4, since by Theorem 4 and Postulate 2

$$(1 \circ a) \circ (b \circ a) = [(1 \circ a) \circ (1 \circ a)] \circ b = 1 \circ b;$$
$$(1 \circ a) \circ (c \circ a) = [(1 \circ a) \circ (1 \circ a)] \circ c = 1 \circ c.$$

(ii) follows from (i) and Theorem 4.1.

THEOREM 6. *If $a, b, c$ are any three elements of $\mathfrak{S}$, the three relations*

$$a \circ b = c, \tag{2}$$

$$b = (1 \circ c) \circ (1 \circ a), \tag{3}$$

$$a = c \circ (1 \circ b) \tag{4}$$

*are all equivalent to one another.*

Note that in accordance with our definitions in §1, (3) and (4) give the first and second inverses of the operation $\circ$ in terms of $\circ$ itself.

(2) implies (3); for if $a \circ b = c$, then by Theorem 4.1,

$$1 \circ c = 1 \circ (a \circ b) = b \circ a.$$

Hence

$$(1 \circ c) \circ (1 \circ a) = (b \circ a) \circ (1 \circ a) = b \circ [(1 \circ a) \circ (1 \circ a)] = b \circ 1 = b,$$

by Postulate 3, Postulate 4 and Theorem 2. Conversely, (3) implies (2); for all the steps in the reasoning above are reversible.

(3) is equivalent to (4); for writing $(1 \circ c)$, $(1 \circ a)$, $b$ for $a, b, c$ in (2), we see from what we have just proved that (3) is equivalent to

$$1 \circ a = (1 \circ b) \circ (1 \circ (1 \circ c)).$$

Hence

$$1 \circ a = (1 \circ b) \circ c = 1 \circ (c \circ (1 \circ b))$$

by Theorem 3 and Postulate 3. Hence by Postulate 2, (3) is equivalent to (4), so that each of (2), (3) and (4) implies the other two.

5. **The operation $\square$.** We shall now study the second inverse of $\circ$ as defined by equation (4) of Theorem 6.

DEFINITION 2. *If $x, y$ are any two elements of $\mathfrak{S}$,*

$$x \square y = y \circ (1 \circ x).$$

FUNDAMENTAL THEOREM. *$\mathfrak{S}$ forms a group with respect to the operation $\square$.*

(i) By Definition 2, Theorem 1 and Postulate 1, if $a$ and $b$ are any elements of $\mathfrak{S}$, $a \square b$ is an element of $\mathfrak{S}$ uniquely determined by $a$ and $b$.

(ii) If $a, b, c$ are any elements of $\mathfrak{S}$, then

$$a \square (b \square c) = (a \square b) \square c.$$

For by Definition 2,

$$(a \square b) \square c = c \circ [1 \circ (b \circ (1 \circ a))];$$

$$a \square (b \square c) = [c \circ (1 \circ b)] \circ (1 \circ a).$$

Now

$$[c \circ (1 \circ b)] \circ (1 \circ a) = c \circ [(1 \circ a) \circ (1 \circ (1 \circ b))] = c \circ [(1 \circ a) \circ b],$$

by Postulate 3, Theorem 2. And by Theorem 4, Theorem 2,

$$c \circ [1 \circ (b \circ (1 \circ a))] = c \circ [(1 \circ (1 \circ (1 \circ a))) \circ b] = c \circ [(1 \circ a) \circ b].$$

(iii) The set contains an element $i$ such that for any element $a$ of the set, $i \square a = a \square i = a$.

For by Definition 2, Theorems 1, 2, 4, $\mathfrak{S}$ contains 1 and

$$1 \square a = a \circ (1 \circ 1) = a \circ 1 = a,$$
$$a \square 1 = 1 \circ (1 \circ a) = a,$$

so that we may take $i = 1$.

(iv) If $a$ is any element of $\mathfrak{S}$, the set also contains an element $a'$ such that $a \square a' = i$.

For by Definition 2, Postulate 2,

$$a \square (1 \circ a) = (1 \circ a) \circ (1 \circ a) = 1.$$

Hence we may take $a' = 1 \circ a$, and the system $\{\mathfrak{S}; \square\}$ satisfies the four postulates for a group.*

Consider the inverses of $\square$ in their relation to the original operation $\circ$.

THEOREM 7. *If* $b \square c = a$, *then* $c = (1 \circ b) \square a = a \circ b$, *and* $b = a \square (1 \circ c) = (1 \circ c) \circ (1 \circ a)$.

This is clear from Definition 2 and Theorem 6.

Thus the first inverse of $\square$ is $\circ$, while the second inverse of $\square$ is, by equation (3), the same as the first inverse of $\circ$.

6. **The operation $\Delta$.** We shall now study the operation defined by equation (3).

DEFINITION 3. *If* $x, y$ *are any two elements of* $\mathfrak{S}$,

$$x \Delta y = (1 \circ x) \circ (1 \circ y).$$

Since as we might expect, the operation $\Delta$ is very similar to the original operation $\circ$, we shall merely state its more important properties. The following four theorems which correspond roughly to Theorems 1 to 6 may all be proved from Definition 3 and the results already given.

THEOREM 8. (i) $a \Delta a = 1$; (ii) $a \Delta 1 = 1 \circ a$; (iii) $1 \Delta a = a$; (iv) $(a \Delta 1) \Delta 1 = a$; (v) $(a \Delta b) \Delta 1 = b \Delta a$; (vi) $1 \circ (a \Delta b) = b \Delta a$.

THEOREM 9. (i) $a \Delta (b \Delta c) = ((b \Delta 1) \Delta a) \Delta c$; (ii) $(a \Delta b) \Delta c = b \Delta ((a \Delta 1) \Delta c)$.

* Speiser, *Die Theorie der Gruppen*, Berlin, 1927, pp. 10–11.

THEOREM 10. *If* $a\Delta b = a\Delta c$, *then* $b=c$; *if* $b\Delta a = c\Delta a$, *then* $b=c$.

THEOREM 11. *If* $b = c\Delta a$, *then* $a \circ b = c$, *and* $b\square c = a$.

For example, Theorem 9 (ii) may be proved as follows.

By Definition 3, $(a\Delta b)\Delta c = [1 \circ (a\Delta b)] \circ (1 \circ c)$. Therefore,

$$(a\Delta b)\Delta c = [b\Delta a] \circ (1 \circ c) = [(1 \circ b) \circ (1 \circ a)] \circ (1 \circ c)$$
$$= (1 \circ b) \circ [(1 \circ c) \circ (1 \circ (1 \circ a))],$$

by Theorem 8 (vi), Definition 3, and Postulate 2. Hence by Theorem 3, Postulate 2, Definition 3,

$$(a\Delta b)\Delta c = (1 \circ b) \circ [(1 \circ c) \circ a] = (1 \circ b) \circ [1 \circ (a \circ (1 \circ c))];$$
$$(a\Delta b)\Delta c = b\Delta[a \circ (1 \circ c)].$$

Putting $b=1$ in this last result,

$$(a\Delta 1)\Delta c = 1\Delta[a \circ (1 \circ c)]$$
$$= a \circ (1 \circ c)$$

by Theorem 8 (iii). Hence

$$(a\Delta b)\Delta c = b\Delta((a\Delta 1)\Delta c).$$

7. **Postulates for the operation Δ.** It remains to give a set of postulates for the operation $\Delta$ which shall be consistent and independent.

THEOREM 12. *The system* $\{\mathfrak{S}; \Delta\}$ *satisfies the following four conditions*:

POSTULATE 1. *If* $a, b$ *are any elements of* $\mathfrak{S}$, $a\Delta b$ *is an element of* $\mathfrak{S}$ *uniquely determined by* $a$ *and* $b$.

POSTULATE 2. *If* $a, b$ *are any elements of* $\mathfrak{S}$, $a\Delta a = b\Delta b$.

POSTULATE 5. *If* $a, b, c$ *are any elements of* $\mathfrak{S}$, *and* 1 *is the element of Theorem* 1 *and Theorem* 8, $(a\Delta b)\Delta c = b\Delta((a\Delta 1)\Delta c)$.

POSTULATE 6. *If* $a, b$ *are any elements of* $\mathfrak{S}$, *and* 1 *is the element of Theorem* 1 *and Theorem* 8, *and if* $a\Delta 1 = b\Delta 1$, *then* $a=b$.

The proof is clear from Theorems 8–11.

Table I of §3 may easily be modified so as to show that these postulates are consistent and independent.

8. **Condition that $\square$ be commutative.** We shall now give a condition that the operation $\square$ be commutative, so that $\{\mathfrak{S}; \square\}$ will form an Abelian group.

THEOREM 13. *A necessary and sufficient condition that the operation* $\square$ *of Definition* 2 *be commutative is that, for every pair of elements* $a, b$ *of* $\mathfrak{S}$, $a\Delta b = b \circ a$.

The condition is necessary; for if $\square$ is commutative, and if $a$, $b$ are any two elements of $\mathfrak{S}$, then

$$a \square b = b \square a. \tag{5}$$

Then by Definition 2,

$$b \circ [b \circ (1 \circ a)] = b \circ [a \square b] = b \circ [b \square a] = b \circ [a \circ (1 \circ b)].$$

But

$$b \circ [b \circ (1 \circ a)] = (b \circ a) \circ b; \quad b \circ [a \circ (1 \circ b)] = 1 \circ a,$$

by Theorem 4, Theorem 3 and Theorem 1. Hence (5) implies that

$$(b \circ a) \circ b = 1 \circ a.$$

By Theorem 6 and Definition 3, this last equation is equivalent to

$$b \circ a = (1 \circ a) \circ (1 \circ b) = a \Delta b.$$

The condition is moreover sufficient, for all the steps in the reasoning just given are reversible.

We close with a table giving the relations between the various operations and their inverses.

TABLE II

| Operation | First Inverse | Second Inverse |
|---|---|---|
| $z = x \circ y$ | $y = z \Delta x$ | $x = y \square z$ |
| $x = y \square z$ | $z = x \circ y$ | $y = z \Delta x$ |
| $y = z \Delta x$ | $x = y \square z$ | $z = x \circ y$ |

I wish to express my thanks to the editors of this journal for some extremely helpful criticisms and suggestions.

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIF.

# ON SEQUENCES OF CONTINUOUS FUNCTIONS HAVING CONTINUOUS LIMITS*

BY

D. C. GILLESPIE AND W. A. HURWITZ

1. **Introduction, notation and terminology.** In order that the limit of a convergent sequence of continuous functions be a continuous function, it is sufficient, but not necessary, that the convergence be uniform; a necessary and sufficient condition is that the convergence be quasi-uniform. Consideration of the theory of summability leads to the observation that it is also sufficient that some sequence into which the given sequence can be transformed by a regular transformation should be uniformly convergent. It may happen that this condition is fulfilled when the original sequence converges non-uniformly. For example, let

$$s_n(x) = \begin{cases} 2^n x, & 0 \leqq x \leqq 1/2^n; \\ 2 - 2^n x, & 1/2^n \leqq x \leqq 1/2^{n-1}; \\ 0, & 1/2^{n-1} \leqq x \leqq 1. \end{cases}$$

Then $s_n(x)$ is continuous, $0 \leqq x \leqq 1$. Furthermore, $s_n(x) \to 0$; for if $x=0$, $s_n(x)=0$ for all $n$, and if $0<x\leqq 1$, $s_n(x)=0$ when $n>1-(\log x)/(\log 2)$. The convergence is non-uniform, since $s_n(1/2^n)=1$. For the first Cesàro mean,

$$\sigma_n(x) = \frac{1}{n}\{s_1(x) + s_2(x) + \cdots + s_n(x)\},$$

we have

$$\sigma_n(x) = \begin{cases} \dfrac{1}{n}(2^{n+1} - 2)x, & 0 \leqq x \leqq 1/2^n; \\ \dfrac{1}{n}(2 - 2x), & 1/2^n \leqq x \leqq 1. \end{cases}$$

It follows that

$$0 \leqq \sigma_n(x) \leqq \frac{2}{n}, \quad 0 \leqq x \leqq 1,$$

and $\sigma_n \to 0$ uniformly.

It is natural to inquire whether every sequence $\{s_n(x)\}$, bounded for all $n$ and $x$, having its elements continuous, and converging to a continuous limit, can be transformed by an appropriately chosen regular transformation

* Presented to the Society, December 27, 1929; received by the editors November 22, 1929.

into a uniformly convergent sequence; the present paper gives an affirmative answer to this question.* The primary object was to establish the result for a bounded closed interval of ordinary one-dimensional euclidean space; but in the course of the proof it became necessary to replace the given interval by a closed subset, so that nothing was lost by starting at once with any bounded closed set. Cast in this form, the work was seen to apply to sets in an abstract metric space; we therefore give all our results in this form. The reader not interested in the general case will find that every step has its usual meaning in a euclidean space of one or more dimensions if the term *compact* is interpreted as *bounded.*

Throughout, we denote by $x$ a point of a metric space. We write $AB$ for the greatest common subset, $A+B$ for the union, of two point sets $A$ and $B$, $\bar{A}$ for the complement (in the whole space) of $A$, and $\subset$ or $\supset$ for the relation of inclusion. If $A$ is a compact closed set and $f(x)$ a function continuous in $A$, then $f(x)$ possesses a maximum in $A$, which we denote by $G(f; A)$.

Let $S$ represent a sequence

$$S: \quad s_1(x),\ s_2(x),\ s_3(x),\ \cdots \tag{1}$$

of functions each continuous in $A$, such that, for some constant $M$,

$$0 \leqq s_n(x) \leqq M, \tag{2}$$

and such that

$$\lim_{n\to\infty} s_n(x) = 0; \tag{3}$$

then we call $S$ an $\mathfrak{S}$-sequence in $A$.

2. **The maximal function.** Let $x$ be a point of $A$. For every sequence† $(x_k)$ of points of $A$ having $x$ as limit and every sequence $(n_k)$ of positive integers becoming infinite, form

$$\limsup_{n\to\infty} s_{n_k}(x_k) = \lambda; \tag{4}$$

the least upper bound of all such numbers $\lambda$ is the value at $x$ of the *maximal function*‡ of $S$ in $A$; it will be denoted by $H(S; A; x)$. At every point of $A$,

* It is evident that such a condition is more readily adapted to practical application than the Arzelà condition of quasi-uniform convergence. For instance, it shows immediately that the sequence of integrals of the elements of a sequence of the stated character is summable to the integral of the limit function.

† It is not assumed that $x_k \neq x$.

‡ Essentially the same concept was used by Osgood, American Journal of Mathematics, vol. 19 (1897), p. 166. Our formulation is identical with that of Hahn, *Theorie der reellen Funktionen*, 1921, p. 231.

$H(S; A; x) \geqq 0$; at an isolated point of $A$, $H(S; A; x) = 0$. For convenience we define $H(S; A; x) = 0$ also at every point of $\bar{A}$.

I. *If $S$ is an $\mathfrak{S}$-sequence in the compact closed set $A$, $H(S; A; x)$ is an upper semi-continuous function. If $B$ is a closed subset of $A$, then $H(S; A; x) \geqq H(S; B; x)$.*

The semi-continuity is a standard result.* The inequality follows at once from the definition.

II. *If $S$ is an $\mathfrak{S}$-sequence in a compact closed set $A$, $h > 0$, and $H(S; A; x) < h$, then*

$$\limsup_{n \to \infty} G(s_n; A) \leqq h. \tag{5}$$

Suppose it were true that

$$\limsup_{n \to \infty} G(s_n; A) > h.$$

Then for a sequence of values $(n_k)$, $G(s_{n_k}; A) > h$. For each $k$ there is a point $x_k$ of $A$ for which $s_{n_k}(x_k) = G(s_{n_k}; A)$; hence

$$s_{n_k}(x_k) > h. \tag{6}$$

The sequence of points $(x_k)$ has at least one limiting point $\xi$, which is a point of $A$. Let the notation be altered so that the indices $k$ refer only to a subsequence $(x_k)$ having $\xi$ as limit. Then (6) holds for all $k$, and

$$\limsup_{k \to \infty} s_{n_k}(x_k) \geqq h. \tag{7}$$

The left side of (7) is one of the values $\lambda$ of (4); thus $H(S; A; \xi) \geqq h$, in contradiction to the hypothesis.

III. *If $S$ is an $\mathfrak{S}$-sequence in a compact closed set $A$ and $h > 0$, then the set $A^1$ of points at which $H(S; A; x) \geqq h$ is a closed proper subset of $A$, nowhere dense in $A$.*

$A^1$ is closed† on account of the upper semi-continuity of $H(S; A; x)$. It is certainly a proper subset of $A$ if it is nowhere dense in $A$. To prove that $A^1$ is nowhere dense in $A$ we show that in each open set $C$ which contains a point of $A$ there is an open set which contains a point of $A$ and no point of $A^1$. Suppose this were untrue for a given open set $C$ containing a point $x_0$ of $A$. Then $C$ itself contains a point of $A^1$; that is, there is a point in $C$ for which $H \geqq h > h/2$. Thus there is an index $n_1$ and a point $x_1$ of $AC$ such that

* Hahn, loc. cit., chapter IV, §1, Theorem IX.

† Hahn, loc. cit., chapter II, §9, Theorem IV.

$s_{n_1}(x_1)>h/2$; and on account of the continuity of $s_{n_1}(x)$ there is an open set $C_1$, contained in $C$ and containing $x_1$, such that $s_{n_1}(x)>h/2$ throughout $AC_1$. As $C_1$ contains a point of $A$, it contains a point of $A^1$, at which $H\geqq h>h/2$; there is an index $n_2>n_1$ and a point $x_2$ of $AC_1$ such that $s_{n_2}(x_2)>h/2$; and there is then an open set $C_2$, contained in $C_1$ and containing $x_2$, such that $s_{n_2}(x>h/2)$ throughout $AC_2$. Continuing thus, we obtain a sequence of open sets, $C_1, C_2, C_3, \cdots$, each contained in the preceding, and such that $s_{n_k}(x_k)>h/2$ throughout $AC_k$. Let $L_k$ denote the set obtained by adjoining to $AC_k$ its limit points. Then each of the sequence of closed sets $L_1, L_2, L_3, \cdots$ contains the following, and $s_{n_k}(x_k)\geqq h/2$ throughout $L_k$. These sets have at least one point in common. At this point $\xi$, $s_{n_k}(\xi)\geqq h/2$ for every $k$. If we let $k\to\infty$, we have $0\geqq h/2$, in contradiction to the hypothesis that $h>0$. We have thus shown that $A^1$ is nowhere dense in $A$.

IV. *If $S$ is an $\mathfrak{S}$-sequence in a compact closed set $A$, $B$ a closed subset of $A$, and $h>0$, and if $A^1$, $B^1$ denote respectively the sets for which $H(S; A; x)\geqq h$, $H(S; B; x)\geqq h$, then $B^1\subset A^1$.*

This follows from I.

3. **$\mathfrak{T}$-transformations.** The transformations used to convert the sequence $S$ into a uniformly convergent sequence will be of a particular type, which we now discuss. Write the transformation

$$T: \qquad \sigma_n = \sum_{k=1}^{\infty} a_{n,k} s_k, \tag{8}$$

which carries the sequence $S$ into the sequence $\Sigma$:

$$\Sigma: \qquad \sigma_1(x), \sigma_2(x), \sigma_3(x), \cdots \tag{9}$$

in the abbreviated form $\Sigma=T(S)$. Such a transformation will be called a *$\mathfrak{T}$-transformation* provided the following conditions are satisfied:

$$a_{n,k}\geqq 0; \quad \sum_{k=1}^{\infty} a_{n,k} = 1; \tag{10}$$

*there exist integers $\mu_n$, $\nu_n$ such that*

$$\mu_1\leqq\nu_1<\mu_2\leqq\nu_2<\mu_3\leqq\nu_3<\cdots \tag{11}$$

*and that*

$$a_{n,k}=0 \quad \textit{unless} \quad \mu_n\leqq k\leqq\nu_n. \tag{12}$$

Obviously (12) makes it possible to write (8) in the form

$$\sigma_n = \sum_{k=\mu_n}^{\nu_n} a_{n,k} s_k,$$

so that each row of the matrix $\|a_{n,k}\|$ contains a finite set of non-zero elements and each column contains at most one such element.

It may be noted that *the identity*

$$I: \qquad \sigma_n = s_n \tag{13}$$

*is a* $\mathfrak{T}$*-transformation, and that the product of two* $\mathfrak{T}$*-transformations is a* $\mathfrak{T}$*-transformation.*

V. *A* $\mathfrak{T}$*-transformation is regular. Furthermore if* $\Sigma = T(S)$*, where* $S$ *is an* $\mathfrak{S}$*-sequence in the compact closed set* $A$*, then*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A).$$

From (10), (11), (12) it is clear that the conditions for regularity are satisfied. To prove the second statement, write

$$\limsup_{n\to\infty} G(s_n; A) = l.$$

For each $\epsilon > 0$ there is an index $K$ such that when $k > K$, $G(s_k; A) < l + \epsilon$.

Therefore when $n > N = \mu_K$,

$$\sigma_n \leqq \sum_{k=\mu_n}^{\nu_n} a_{n,k} G(s_k; A) < (l + \epsilon) \sum_{k=\mu_n}^{\nu_n} a_{n,k} = l + \epsilon$$

throughout $A$, so that

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq l + \epsilon.$$

Since this result holds for arbitrary $\epsilon > 0$, it follows that

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq l.$$

A $\mathfrak{T}$-transformation, as defined, is not of the familiar form which possesses a triangular matrix; it is, however, equivalent to such a transformation. Let

$$b_{n,k} = \begin{cases} a_{m,k} \text{ if } \nu_m \leqq n < \nu_{m+1},\ \mu_m \leqq k \leqq \nu_m; \\ 0 \quad \text{otherwise.} \end{cases} \tag{14}$$

Then the transformation

$$\tau_n = \sum_{k=1}^{n} b_{n,k} s_k \tag{15}$$

has a triangular matrix, and is related to (8) as follows:

$$\tau_n = \begin{cases} 0, & n < \nu_1; \\ \sigma_m, & \nu_m \leqq n < \nu_{m+1}. \end{cases} \tag{16}$$

Thus the sequence $(\tau_n)$ is obtained (for $n \geqq \nu_1$) by appropriate repetition of the terms of $(\sigma_n)$. If either of these sequences converges, the other converges to the same value; if either converges uniformly, the other converges uniformly.

4. **Application of $\mathfrak{T}$-transformations to $\mathfrak{S}$-sequences.** We now prove several lemmas to facilitate the work of §§6, 7.

VI. *Let $h \geqq 0$; let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$. If for each $q > h$ there is a $\mathfrak{T}$-transformation $\Sigma^{(q)} = T_q(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n^{(q)}; A) \leqq q, \tag{17}$$

*then there is a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq h. \tag{18}$$

Let $q, q_1, q_2, \cdots$ be a decreasing sequence of numbers having the limit $h$. Since $\limsup G(\sigma_n^{(q_1)}; A) \leqq q_1 < q$, we can choose $n_1$ so that $G(\sigma_{n_1}^{(q_1)}; A) < q$. Again since $\limsup G(\sigma_n^{(q_2)}; A) \leqq q_2 < q_1$, it follows that for all sufficiently great $n$, $G(\sigma_n^{(q_2)}; A) < q_1$. Choose $n_2$ so great as to satisfy this condition and also so great that $\sigma_{n_2}^{(q_2)}$ involves only elements of $S$ of greater subscript than those found in $\sigma_{n_1}^{(q_1)}$; this is possible since each $T_q$ is a $\mathfrak{T}$-transformation. Continuing thus, choose generally $\sigma_{n_k}^{(q_k)}$ so that

$$G(\sigma_{n_k}^{(q_k)}; A) < q_{k-1} \tag{19}$$

and so that $\sigma_{n_k}^{(q_k)}$ contains only elements of $S$ of subscript greater than those found in $\sigma_{n_{k-1}}^{(q_{k-1})}$. Now define

$$T: \qquad \sigma_k = \sigma_{n_k}^{(q_k)}.$$

This is a $\mathfrak{T}$-transformation. Also, by (19),

$$G(\sigma_k; A) < q_{k-1},$$

and (18) follows at once.

In VI replace $h$ by 0, $q$ by $h$; we state this special case:

VII. *Let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$. If for each $h > 0$ there is a $\mathfrak{T}$-transformation $\Sigma^{(h)} = T_h(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n^{(h)}; A) \leqq h,$$

*then there is a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that $\sigma_n \to 0$ uniformly in $A$.*

VIII. *Let $h>0$; let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$, $B$ a closed subset of $A$. If*

$$\limsup_{n\to\infty} G(s_n; B) < h, \tag{20}$$

*and if for each neighborhood* $C$ of $B$ in $A$ there exists a $\mathfrak{T}$-transformation $\Sigma^{(C)} = T^{(C)}(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n^{(C)}; A\overline{C}) < h, \tag{21}$$

*then there exists a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq h. \tag{22}$$

Denote an index $n$ chosen at random (for example, $n=1$) by $m_{1,1}$ and define

$$\sigma_1 = s_{m_{1,1}}. \tag{23}$$

This will be the first formula in the definition of $T$.

On account of (20), for all sufficiently great $n$, $s_n<h$ throughout $B$. Choose $m_{2,1}>m_{1,1}$ so that $s_{m_{2,1}} < h$ throughout $B$. Since $s_{m_{2,1}}$ is continuous, there is a neighborhood $C_{2,1}$ of $B$ in $A$ such that† $s_{m_{2,1}}<h$ throughout $C_{2,1}$. Form the transformation $T^{(C_{2,1})}$; by (21), for all sufficiently great $n$, $\sigma_n^{(C_{2,1})}<h$ in $A\overline{C}_{2,1}$. Choose $m_{2,2}$ so that $\sigma_{m_{2,2}}^{(C_{2,1})}<h$ in $A\overline{C}_{2,1}$, and so that $\sigma_{m_{2,2}}^{(C_{2,1})}$ contains only elements of $S$ of subscript greater than $m_{1,1}$. Define

$$\sigma_2 = \tfrac{1}{2}\left\{ s_{m_{2,1}} + \sigma_{m_{2,2}}^{(C_{2,1})} \right\}. \tag{24}$$

We see that $\sigma_2$ involves elements of $S$ of a finite set of indices of which the least is greater than $m_{1,1}$.

We continue the definition by induction. Suppose that $\sigma_1, \sigma_2, \cdots, \sigma_{k-1}$ have been written so that each contains terms of $S$ having a finite set of indices, of which the least in any row is greater than the greatest in the preceding row; we show how to form $\sigma_k$. Call $\nu$ the greatest subscript of an element of $S$ appearing in $\sigma_{k-1}$. As before choose $m_{k,1}>\nu$ and such that $s_{m_{k,1}}<h$ throughout $B$; then choose a neighborhood $C_{k,1}$ of $B$ in $A$ such that $s_{m_{k,1}}<h$ throughout $C_{k,1}$. Form $T^{(C_{k,1})}$; choose $m_{k,2}>\nu$ so that $\sigma_{m_{k,2}}^{(C_{k,1})}<h$ in $A\overline{C}_{k,1}$ and also $\sigma_{m_{k,2}}^{(C_{k,1})}<h$ in $B$; then choose a neighborhood $C_{k,2}$ of $B$ in $C_{k,1}$ such that $\sigma_{m_{k,2}}^{(C_{k,1})}<h$ in $C_{k,2}$. Next form $T^{(C_{k,2})}$; choose $m_{k,3}>\nu$ so that

* By a neighborhood of $B$ in $A$ we mean a subset of $A$ open in $A$ and containing $B$.

† For each point $b$ in $B$ a neighborhood of $b$ in $A$ exists throughout which $s_{m_{2,1}}<h$; the union of all such point-neighborhoods will be a possible $C_{2,1}$.

$\sigma_{m_{k,3}}^{(C_{k,2})} < h$ in $A\overline{C}_{k,2}$ and also $\sigma_{m_{k,3}}^{(C_{k,2})} < h$ in $B$; then choose a neighborhood $C_{k,3}$ of $B$ in $C_{k,2}$ such that $\sigma_{m_{k,3}}^{(C_{k,2})} < h$ in $C_{k,3}$. Continue thus, choosing the indices, neighborhoods, and transformations:

$$m_{k,1}, C_{k,1}, T^{(C_{k,1})}; m_{k,2}, C_{k,2}, T^{(C_{k,2})}; \cdots;$$
$$m_{k,k-1}, C_{k,k-1}, T^{(C_{k,k-1})}; m_{k,k};$$

so that

$$A \supset C_{k,1} \supset C_{k,2} \supset \cdots \supset C_{k,k-1}; \tag{25}$$

$$\begin{cases} s_{m_{k,1}} < h \quad \text{in} \quad C_{k,1}; & \\ \sigma_{m_{k,p+1}}^{(C_{k,p})} < h \quad \text{in} \quad A\overline{C}_{k,p} & (p = 1,2, \cdots, k-1); \\ \sigma_{m_{k,p+1}}^{(C_{k,p})} < h \quad \text{in} \quad C_{k,p+1} & (p = 1,2, \cdots, k-2). \end{cases} \tag{26}$$

The set $A$ can be decomposed into the following parts having no points in common:

$$A = A\overline{C}_{k,1} + C_{k,1}\overline{C}_{k,2} + C_{k,2}\overline{C}_{k,3} + \cdots + C_{k,k-2}\overline{C}_{k,k-1} + C_{k,k-1}.$$

In $A\overline{C}_{k,1}$ (which is contained in $A\overline{C}_{k,p}$ for $p \leqq k-1$), by (2) and (26),

$$s_{m_{k,1}} \leqq M; \quad \sigma_{m_{k,p}}^{(C_{k,p-1})} < h \text{ for } p > 1.$$

In $C_{k,r-1}\overline{C}_{k,r}$ (which is contained in $A\overline{C}_{k,p}$ for $p \geqq r$ and in $C_{k,p}$ for $p \leqq r-1$), by (2), (26), and (10),

$$s_{m_{k,1}} < h; \quad \sigma_{m_{k,p}}^{(C_{k,p-1})} < h \text{ for } p \neq r; \quad \sigma_{m_{k,r}}^{(C_{k,r-1})} \leqq M.$$

In $C_{k,k-1}$ (which is contained in $C_{k,p}$ for $p \leqq k-1$),

$$s_{m_{k,1}} < h; \quad \sigma_{m_{k,p}}^{(C_{k,p-1})} < h \text{ for } p \neq k; \sigma_{m_{k,k}}^{(C_{k,k-1})} \leqq M.$$

Thus throughout $A$

$$s_{m_{k,1}} + \sigma_{m_{k,2}}^{(C_{k,1})} + \sigma_{m_{k,3}}^{(C_{k,2})} + \cdots + \sigma_{m_{k,k}}^{(C_{k,k-1})} < (k-1)h + M. \tag{27}$$

We now define the $k$th row of the transformation $T$ as follows:

$$\sigma_k = \frac{1}{k}\left\{ s_{m_{k,1}} + \sigma_{m_{k,2}}^{(C_{k,1})} + \sigma_{m_{k,3}}^{(C_{k,2})} + \cdots + \sigma_{m_{k,k}}^{(C_{k,k-1})} \right\}. \tag{28}$$

The transformation which has thus been defined is a $\mathfrak{T}$-transformation, since (10), (11), (12) are easily seen to hold. Also, from (27) and (28) we have throughout $A$

$$\sigma_k < \frac{1}{k}\{(k-1)h + M\} = h + \frac{M-h}{k};$$

thus

$$G(\sigma_n; A) < h + \frac{M-h}{n},$$

and (22) follows at once.

IX. *Let $h>0$; let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$, $B$ a closed subset of $A$. If*

$$\limsup_{n\to\infty} G(s_n; B) \leqq h,$$

*and if for each neighborhood $C$ of $B$ in $A$ there is a $\mathfrak{T}$-transformation $\Sigma^{(C)} = T^{(C)}(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n{}^{(C)}; A\overline{C}) \leqq h,$$

*then there exists a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq h.$$

Let $q>h$. Then $q$ satisfies the conditions imposed on $h$ in VIII; thus for each $q>h$, VIII furnishes a $\mathfrak{T}$-transformation $\Sigma^{(q)} = T_q(S)$ such that

$$\limsup_{n\to\infty} G(\sigma_n^{(q)}; A) \leqq q.$$

The conclusion of the present theorem then follows from VI.

5. **Definition and properties of $h$-sets and $h$-order.** Let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$ and let $h>0$. We define a set of sets $A^\alpha$, where $\alpha$ is any Cantor ordinal of the first or second class, by the following scheme of induction:

(1) $A^0 = A$;

(2) if $\alpha$ is not a limiting ordinal, then $A^\alpha$ is the set of points at which $H(S; A^{\alpha-1}; x) \geqq h$;

(3) if $\alpha$ is a limiting ordinal, then $A^\alpha$ is the greatest common subset of all $A^\rho$ for $\rho<\alpha$.

The sets thus defined we call the *$h$-sets generated by $(S; A)$*. By the process of definition the set of $h$-sets is normally ordered. Several properties will be noted.

X. *If $B$ is a closed subset of $A$ and if $B^\alpha$ denote the $h$-sets for $(S; B)$, then $B^\alpha \subset A^\alpha$.*

The statement holds, by the hypothesis, when $\alpha=0$. We show that if it holds for each ordinal $\rho<\alpha$, it holds also for $\alpha$. Suppose that whenever $\rho<\alpha$, $B^\rho \subset A^\rho$. If $\alpha$ is not a limiting ordinal, $B^{\alpha-1} \subset A^{\alpha-1}$, and by IV, $B^\alpha \subset A^\alpha$. If $\alpha$ is a limiting ordinal, $B^\alpha$ is the greatest common subset of all $B^\rho$ for $\rho<\alpha$; each such $B^\rho \subset A^\rho$; hence $B^\alpha$ is contained in the greatest common subset of all $A^\rho$ for $\rho<\alpha$, which is $A^\alpha$.

XI. *For $(S\,;A)$ each h-set is a closed subset of each preceding h-set; each h-set is a proper subset of each preceding non-empty h-set.*

That $A^1$ is a closed proper subset of $A^0$ is stated in III. Suppose that for $\rho<\alpha$, $A^\rho$ is a closed subset of every earlier $h$-set and a proper subset of each earlier non-empty $h$-set. If $\alpha$ is not a limiting ordinal it follows from III, with $A$ replaced by $A^{\alpha-1}$, that $A^\alpha$ is a closed proper subset of $A^{\alpha-1}$ (and hence of all earlier $h$-sets) unless $A^{\alpha-1}$ is empty; if $A^{\alpha-1}$ is empty, $A^\alpha$ will also be empty, since $H(S;A^{\alpha-1};x)$ will then be zero throughout the space. If $\alpha$ is a limiting ordinal, and no preceding $h$-set is empty, let $\alpha_n\to\alpha$; then $A^\alpha$ is the greatest common subset of all $A^{\alpha_n}$, and any ordinal $\beta<\alpha$ is followed by some $\alpha_n$, so that $A^\alpha$ is a subset of $A^{\alpha_n}$, which is a proper subset of $A^\beta$; if $\alpha$ is a limiting ordinal and some preceding $h$-set is empty, $A^\alpha$ is empty.

XII. *For $(S\,;A)$ there is a non-limiting ordinal $\gamma$ such that $A^\gamma$ and all following h-sets are empty, while all preceding h-sets are non-empty.*

If $A^\alpha$ were non-empty for every $\alpha$, the set of $h$-sets would constitute a normally ordered set of closed sets such that each is a proper subset of all its predecessors; such a set of sets must be countable;* but the set of ordinals of first and second classes is not countable,† so that we should have a contradiction. Let $\gamma$ be the first ordinal for which $A^\gamma$ is empty; then $\gamma$ is not a limiting ordinal. For if $\gamma$ were a limiting ordinal, let $\gamma_n\to\gamma$; then $A^\gamma$ would be the greatest common subset of the descending sequence of non-empty closed sets $A^{\gamma_n}$ and hence could not be empty.‡

Since $\gamma$, as given by XII, is not a limiting ordinal, we may write $\alpha=\gamma-1$. Then $A^\alpha$ and all preceding $h$-sets are non-empty, while all following $h$-sets are empty. We define the *h-order of* $(S;A)$ as the greatest ordinal $\alpha$ for which $A^\alpha$ is not empty.

XIII. *If $\alpha$ is the h-order of $(S\,;A)$ then*

$$\limsup_{n\to\infty} G(s_n;A^\alpha) \leqq h.$$

---

* Hahn, loc. cit., chapter I, §7, Theorem XII.

† Hahn, loc. cit., Introduction, §4, Theorem XIV.

‡ Hahn, loc. cit., chapter I, §2, Theorem VIII.

Since $A^{\alpha+1}$ (which is the set of points at which $H(S; A^\alpha; x) \geqq h$) is empty, $H(S; A^\alpha; x) < h$; the conclusion follows from II with $A$ replaced by $A^\alpha$.

**6. Transformation of an $\mathfrak{S}$-sequence into a uniformly convergent sequence.** In this section we prove two theorems which imply the principal result given in the next section.

XIV. *Let $h>0$. If $S$ is an $\mathfrak{S}$-sequence in the compact closed set $A$, there exists a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq h. \tag{29}$$

We prove this by ordinal induction on $\alpha$, the $h$-order of $(S; A)$. If $\alpha = 0$, then (29) holds, by XIII, since the identity is a $\mathfrak{T}$-transformation. It will be shown that if the conclusion holds for every closed compact set $Q$ such that the $h$-order of $(S; Q)$ is less than $\alpha$, it holds for the set $A$ such that the $h$-order of $(S; A)$ is $\alpha$.

Let $C$ be a neighborhood of $A^\alpha$ in $A$. Since $A^\alpha \subset C$, $A^\alpha \overline{C}$ is empty. Form $Q = A\overline{C}$ and let $Q^\rho$ represent the $h$-sets generated by $(S; Q)$. Then $Q^\rho \subset A^\rho$, by X, and $Q^\rho \subset Q \subset \overline{C}$; hence $Q^\rho \subset A^\rho \overline{C}$. Thus $Q^\alpha$ is empty, and the $h$-order of $(S; Q)$ is some ordinal $\beta < \alpha$. The hypothesis of our theorem must then hold with $A$ replaced by $Q$ and $\alpha$ by $\beta$; the conclusion yields a $\mathfrak{T}$-transformation $\Sigma' = T'(S)$ such that

$$\limsup_{n\to\infty} G(\sigma_n'; Q) \leqq h.$$

The transformation thus obtained depends on $Q$, and hence on the choice of $C$; indicating this by proper notation and replacing $Q$ by $A\overline{C}$, we see that for each neighborhood $C$ of $A^\alpha$ in $A$ we have found a $\mathfrak{T}$-transformation $\Sigma^{(C)} = T^{(C)}(S)$ such that

$$\limsup_{n\to\infty} G(\sigma_n^{(C)}; A\overline{C}) \leqq h.$$

But by XIII,

$$\limsup_{n\to\infty} G(s_n; A^\alpha) \leqq h.$$

The hypotheses of IX are then satisfied with $B$ replaced by $A^\alpha$; the conclusion of IX furnishes the conclusion of our theorem.

XV. *Let $S$ be an $\mathfrak{S}$-sequence in the compact closed set $A$. Then there exists a $\mathfrak{T}$-transformation $\Sigma = T(S)$ such that $\sigma_n \to 0$ uniformly in $A$.*

For each $h>0$ there exists by XIV a $\mathfrak{T}$-transformation $\Sigma^{(h)} = T_h(S)$ such that

$$\limsup_{n\to\infty} G(\sigma_n^{(k)}; A) \leqq h;$$

the theorem follows from VII.

**7. The principal theorem.** We easily deduce from XV the result mentioned in §1.

XVI. *Let $A$ be a compact closed set; let the sequence*

$$s_1(x), s_2(x), s_3(x), \cdots$$

*have the following properties*:

(i) *for each $n$, $s_n(x)$ is continuous in $A$;*

(ii) *for each $x$ in $A$, $s_n(x)\to s(x)$;*

(iii) *$s(x)$ is continuous in $A$;*

(iv) *there is a constant $M$ such that for all $x$ in $A$ and all $n$, $|s_n(x)|\leqq M$. Then there exists a $\mathfrak{T}$-transformation $\Sigma=T(S)$ such that $\sigma_n(x)\to s(x)$ uniformly in $A$.*

For all $x$ in $A$, $|s(x)|\leqq M$, by (i), (iv), and hence $|s_n(x)-s(x)|\leqq 2M$. Call $s_n'(x)=|s_n(x)-s(x)|$. Then

$$S': \quad s_1'(x), s_2'(x), s_3'(x), \cdots$$

is an $\mathfrak{S}$-sequence in $A$. By XV there is a $\mathfrak{T}$-transformation which, expressed in terms of $S'$, gives $\Sigma'=T(S')$ such that $\sigma_n'\to 0$ uniformly in $A$. Let $\Sigma=T(S)$. Since no coefficient in a $\mathfrak{T}$-transformation is negative and the sum of the coefficients in any row is 1,

$$\begin{aligned}|\sigma_n - s| &= |(\Sigma a_{n,k} s_k) - s| = |\Sigma a_{n,k}(s_k - s)| \\ &\leqq \Sigma a_{n,k}|s_k - s| = \Sigma a_{n,k} s_k' = \sigma_n',\end{aligned}$$

and $\sigma_n\to s$ uniformly in $A$.

**8. Limitation on extension of results.** As has been seen, a necessary and sufficient condition that a sequence satisfying (i), (ii), and (iv) of XVI have a continuous limit is that it be uniformly summable by a properly chosen regular transformation. The range of applicability of this result differs from the range of the Arzelà condition only by the assumption of condition (iv). How far, if at all, this condition may be relaxed, remains an open question; it is at any rate possible to show that too rapid increase in the violence of oscillation of the elements of the sequence may entirely prevent uniform summability.

Denote by $J$ the interval $0\leqq x\leqq 1$ and by $J_k$ the interval $1/2^k\leqq x\leqq 1/2^{k-1}$. For each $n$ let* $s_n(x)=0$ in $J$ except in $J_n$; let $s_n(x)=0$ at the ends of $J_n$,

* For example, $s_n$ might be defined by an isosceles triangle of height $M_n$ over $J_n$ and have the value 0 elsewhere in $J$.

have a maximum $M_n \geqq M > 0$ at a point $x_n$ of $J_n$, and be continuous and non-negative in $J_n$. It is clear that $s_n(x) \to 0$ *non-uniformly in* $J$. We show further:

*If the series*

$$\frac{1}{M_1} + \frac{1}{M_2} + \frac{1}{M_3} + \cdots \tag{30}$$

*converges, then there is no regular transformation of the form*

$$\sigma_n(x) = \sum_{k=1}^{\infty} a_{n,k} s_k(x)$$

*such that* $\sigma_n(x) \to 0$ *uniformly in* $J$.

Call $R$ the value of the series (30). We have $\sigma_n(x) = a_{n,k} s_k(x)$ in $J_k$. If $\sigma_n(x) \to 0$ uniformly in $J$, then for each $\epsilon > 0$ there exists $N$ independent of $x$ such that when $n > N$, $|\sigma_n(x)| < \epsilon$ throughout $J$. Thus for $n > N$, $|a_{n,k}| s_k(x) < \epsilon$ in $J_k$. In particular (since $x_k$ is in $J_k$) for $n > N$, $|a_{n,k}| M_k < \epsilon$; that is

$$|a_{n,k}| < \frac{\epsilon}{M_k} \text{ for } n > N; \quad k = 1, 2, 3, \cdots.$$

Hence for $n > N$,

$$\left| \sum_{k=1}^{\infty} a_{n,k} \right| \leqq \sum_{k=1}^{\infty} |a_{n,k}| < \sum_{k=1}^{\infty} \frac{\epsilon}{M_k} = R\epsilon.$$

Therefore

$$\lim_{n \to \infty} \sum_{k=1}^{\infty} a_{n,k} = 0.$$

But for any regular transformation this limit must be 1, and we have a contradiction.

The $\mathfrak{T}$-transformations which we have used in this paper form a class obviously well adapted to the proof of the theorem. It might be asked whether some more special class of transformations of familiar type could be used for the same purpose. Some light is thrown on this question by a simple example of an $\mathfrak{S}$-sequence which cannot be transformed into a uniformly convergent sequence by any Cesâro mean. Let

$$s_n(x) = \begin{cases} nx, & 0 \leqq x \leqq 1/n; \\ n - n(n-1)x, & 1/n \leqq x \leqq 1/(n-1); \\ 0, & 1/(n-1) \leqq x \leqq 1. \end{cases}$$

It is clear that $s_n(x)$ is continuous and has the limit 0 in the interval $0 \leqq x \leqq 1$,

and that $0 \leqq s_n(x) \leqq 1$. Form the Cesàro mean of order $r$:

$$\sigma_n(x) = \frac{\sum_{k=1}^{n} \binom{n-k+r-1}{r-1} s_k(x)}{\binom{n+r-1}{r}}.$$

If $\sigma_n(x) \to 0$ uniformly, then in particular, $\sigma_n(1/n) \to 0$. Since $s_k(1/n) = k/n$ when $k \leqq n$,

$$\sigma_n\left(\frac{1}{n}\right) = \frac{\sum_{k=1}^{n} \binom{n-k+r-1}{r-1} \frac{k}{n}}{\binom{n+r-1}{r}}. \tag{31}$$

To evaluate the numerator, we note that the power series

$$\sum_{k=1}^{\infty} k t^k = t(1-t)^{-2}, \tag{32}$$

$$\sum_{p=0}^{\infty} \binom{r+p-1}{r-1} t^p = (1-t)^{-r}, \tag{33}$$

$$\sum_{n=1}^{\infty} \binom{n+r}{r+1} t^n = t(1-t)^{-r-2} \tag{34}$$

are all absolutely convergent for $|t| < 1$. If we compare the coefficient of $t^n$ in the Cauchy product of (32), (33) with that in (34), we see that

$$\sum_{k=1}^{n} k \binom{n-k+r-1}{r-1} = \binom{n+r}{r+1}.$$

Hence, by (31),

$$\sigma_n\left(\frac{1}{n}\right) = \frac{\frac{1}{n}\binom{n+r}{r+1}}{\binom{n+r-1}{r}} = \frac{n+r}{n} \frac{1}{r+1},$$

and

$$\lim_{n \to \infty} \sigma_n\left(\frac{1}{n}\right) = \frac{1}{r+1} > 0. \tag{35}$$

It is well known that the Euler-Abel power series transformation of a sequence into a function includes all the Cesàro means. Since the limit (35) is more nearly zero the higher the order $r$, it might be hoped that the Euler-

Abel transformation would yield uniformity. However, this is not the case. Write

$$\sigma(x,t) = (1-t)\sum_{k=1}^{\infty} s_k(x)t^{k-1}.$$

If

$$\lim_{t\to 1-}\sigma(x,t) = 0$$

uniformly for all $x$ in the interval, then in particular

$$\lim_{n\to\infty}\sigma\left(\frac{1}{n},\ 1-\frac{1}{n}\right) = 0.$$

But

$$\begin{aligned}\sigma\left(\frac{1}{n},t\right) &= (1-t)\sum_{k=1}^{n}\frac{k}{n}t^{k-1}\\ &= \frac{1-t}{n}\frac{d}{dt}\sum_{k=0}^{n}t^k\\ &= \frac{1-t}{n}\frac{d}{dt}\frac{1-t^{n+1}}{1-t}\\ &= \frac{1+t^n(nt-n-1)}{n(1-t)};\end{aligned}$$

therefore

$$\sigma\left(\frac{1}{n},\ 1-\frac{1}{n}\right) = 1-2\left(1-\frac{1}{n}\right)^n$$

and

$$\lim_{n\to\infty}\sigma\left(\frac{1}{n},\ 1-\frac{1}{n}\right) = 1-\frac{2}{e} > 0.$$

The Borel transformation also fails to bring about uniform convergence. Write

$$\sigma(x,t) = e^{-t}\sum_{k=1}^{\infty}s_k(x)\frac{t^{k-1}}{(k-1)!}.$$

If

$$\lim_{t\to+\infty}\sigma(x,t) = 0$$

uniformly for all $x$ in the interval, then in particular

$$\lim_{n\to\infty}\sigma\left(\frac{1}{n},\ n\right) = 0.$$

We shall prove that, instead,

$$\lim_{n\to\infty} \sigma\left(\frac{1}{n},\ n\right) = \frac{1}{2}. \tag{36}$$

We have

$$\begin{aligned}
\sigma\left(\frac{1}{n},\ t\right) &= e^{-t}\sum_{k=1}^{n} \frac{k}{n}\frac{t^{k-1}}{(k-1)!} \\
&= e^{-t}\sum_{k=1}^{n} \frac{k-1}{n}\frac{t^{k-1}}{(k-1)!} + e^{-t}\sum_{k=1}^{n} \frac{1}{n}\frac{t^{k-1}}{(k-1)!} \\
&= e^{-t}\sum_{k=1}^{n-1} \frac{1}{n}\frac{t^{k}}{(k-1)!} + e^{-t}\sum_{k=1}^{n} \frac{1}{n}\frac{t^{k-1}}{(k-1)!} \\
&= \frac{t+1}{n}e^{-t}\sum_{k=1}^{n} \frac{t^{k-1}}{(k-1)!} - \frac{e^{-t}t^{n}}{n!};
\end{aligned}$$

hence

$$\sigma\left(\frac{1}{n},\ n\right) = \left(1+\frac{1}{n}\right)e^{-n}\sum_{k=1}^{n} \frac{n^{k-1}}{(k-1)!} - \frac{e^{-n}n^{n}}{n!}.$$

By Stirling's theorem, the second term on the right has the limit zero. It thus suffices to show that

$$\lim_{n\to\infty} e^{-n}\sum_{k=1}^{n} \frac{n^{k-1}}{(k-1)!} = \frac{1}{2}$$

or that

$$\lim_{n\to\infty} e^{-n}\sum_{k=1}^{n+1} \frac{n^{k-1}}{(k-1)!} = \frac{1}{2}, \tag{37}$$

since the added term, again by Stirling's theorem, has the limit zero. From the identity

$$e^{-z}\frac{z^{n}}{n!} = \frac{d}{dz}\left\{-e^{-z}\sum_{k=1}^{n+1} \frac{z^{k-1}}{(k-1)!}\right\}$$

we find by integration from 0 to $n$ that

$$\int_{0}^{n} e^{-z}\frac{z^{n}}{n!}dz = 1 - e^{-n}\sum_{k=1}^{n+1} \frac{n^{k-1}}{(k-1)!};$$

thus (37) is equivalent to

$$\lim_{n\to\infty} \int_{0}^{n} e^{-z}\frac{z^{n}}{n!}dz = \frac{1}{2}.$$

This is a standard result.*

It can also be shown that no linear transformation with triangular matrix which is totally regular and permutable with the first Cesàro mean can yield in the present example a uniformly convergent sequence.

Nevertheless, so simple a totally regular transformation as the following will suffice:

$$\sigma_n(x) = \frac{\frac{s_1(x)}{1} + \frac{s_2(x)}{2} + \cdots + \frac{s_n(x)}{n}}{1 + \frac{1}{2} + \cdots + \frac{1}{n}}. \tag{38}$$

In fact, denoting by $D_n$ the denominator of (38), we have

$$\sigma_n(x) = \begin{cases} nx/D_n, & 0 \leqq x \leqq 1/n, \\ 1/D_n, & 1/n \leqq x \leqq 1. \end{cases}$$

Thus

$$0 \leqq \sigma_n(x) \leqq \frac{1}{D_n}, \quad 0 \leqq x \leqq 1;$$

and $\sigma_n(x) \to 0$ uniformly.

* For instance it is a special case of the formula given by Pólya-Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. I, Part 2, Problem 210, with $\alpha=\beta=0$.

CORNELL UNIVERSITY,
ITHACA, N. Y.

# SERIES ASSOCIATED WITH CERTAIN IRREGULAR THIRD-ORDER BOUNDARY VALUE PROBLEMS*

BY

LEWIS E. WARD

1. **Introduction.** In a previous paper† the writer discussed the characteristic functions of the differential system

$$u''' + \rho^3 u = 0,\ u(\pi) = u(-\omega_2\pi) = u(-\omega_3\pi) = 0,$$

where $\omega_2 = e^{\pi i/3}$ and $\omega_3 = e^{-\pi i/3}$, and showed that these characteristic functions fall naturally into three classes, denoted by $u_{1k}(x)$, $u_{2k}(x)$, and $u_{3k}(x)$. It was proved that if $f(x)$ is a function which can be written in the form $\phi_1(x^3)+x\phi_2(x^3)+x^2\phi_3(x^3)$, where each $\phi$ represents a convergent power series in $x^3$ and is such that $\phi_1$, $\phi_2$, and $\phi_3$ have continuous second, third, and fourth derivatives respectively in the interval $0 \leqq x \leqq \pi$,‡ then $f(x)$ can be expanded in a series of these characteristic functions

$$\sum_{k=1}^{\infty} [a_{1k}u_{1k}(x) + a_{2k}u_{2k}(x) + a_{3k}u_{3k}(x)], \tag{1}$$

the series converging uniformly to $f(x)$ in the interior and on the boundary of some equilateral triangle $T_1$ centered at $x=0$, having one vertex on the segment from 0 to $\pi$, and lying inside the smallest of the circles of convergence of the power series for the $\phi$'s.

The purposes of this paper are to consider the uniqueness of this expansion, to extend the region of uniform convergence, to consider the differentiation and the integration of the series, to prove certain relations involving the coefficients, and to show how the coefficients may be obtained by rendering stationary certain integrals.

2. **Uniqueness of the expansion.** The coefficients in the expansion (1),

* Various parts of this paper were presented to the Society on April 7, 1928, March 29, 1929, and August 29, 1929; the paper was received by the editors October 23, 1929.

† *Some third-order irregular boundary value problems*, these Transactions, vol. 29 (1927), pp. 716–745.

‡ Ward, loc. cit., Theorem 1, p. 742. The continuity of the second derivatives only of $\phi_2$ and $\phi_3$ is sufficient in the present case. The continuity of the higher derivatives was demanded in the theorem referred to because that theorem is based on Theorem 2, p. 730 of the same paper, in the proof of which this continuity was useful in connection with the formal relations between the $\phi$'s of the latter theorem. There need be no formal relations between the $\phi$'s of Theorem 1, p. 742, nor between the $\phi$'s of the present paper.

called the formal coefficients, are obtained from certain orthogonality relations, as is customary in expansion problems associated with differential systems. Explicit formulas for them are

$$a_{ik} = \frac{1}{\pi\delta_{i-1}(\rho_{ik}\pi)} \int_0^\pi x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx \qquad (i = 1, 2, 3) \tag{2}$$

where

$$\delta_j(t) = e^{\omega_1 t} + (-\omega_3)^{j-1}e^{\omega_2 t} + (-\omega_2)^{j-1}e^{\omega_3 t}, \quad \omega_1 = -1, \text{ and } \delta_j(t) = \delta_{j+3}(t). *$$

The $u_{ik}(x)$ of the introduction are given by $u_{ik}(x) = \delta_i(\rho_{ik}x)$.

It was shown in the paper referred to that each of the expansions

$$\sum_{k=1}^{\infty} a_{ik}u_{ik}(x) = x^{i-1}\phi_i(x^3) \qquad (i = 1,2,3) \tag{3}$$

is valid in $T_1$. Hence, taking three special cases,

$$1 = \sum_{k=1}^{\infty} b_{1k}u_{1k}(x), \qquad x = \sum_{k=1}^{\infty} b_{2k}u_{2k}(x), \qquad x^2 = \sum_{k=1}^{\infty} b_{3k}u_{3k}(x). \tag{4}$$

It is easily shown by straightforward computation that no one of the $b$'s is zero, and that the triangle of uniform convergence $T_2$ of each of these series can have one vertex anywhere we choose to place it on the interval $0 < x < \pi$.†

Since the terms of each of series (3) are analytic functions, the series may be differentiated term by term as many times as we like. Differentiating three times and making use of the fact that $u_{ik}''' = -\rho_{ik}^3 u_{ik}(x)$, we obtain

$$0 = \sum_{k=1}^{\infty} b_{ik}\rho_{ik}^3 u_{ik}(x) \qquad (i = 1,2,3). \tag{5}$$

Each of these series converges uniformly in and on $T_2$. Adding series (5) for $i = 1$, 2, and 3 to series (1) produces a series

$$\sum_{k=1}^{\infty} [c_{1k}u_{1k}(x) + c_{2k}u_{2k}(x) + c_{3k}u_{3k}(x)]$$

which converges uniformly to $f(x)$ in and on the smaller of $T_1$ and $T_2$, and has coefficients different from the formal coefficients.

Instead of expanding 1, $x$, and $x^2$, as in (4), we may expand $p_1(x^3)$, $xp_2(x^3)$, and $x^2p_3(x^3)$, where the $p$'s are polynomials in $x^3$, and, by differ-

* See p. 720 of my former paper, already referred to, for other formulas for the $\delta$ functions, and p. 741 for the $\rho_{ik}$.

† We shall use the notation $T_i$ to represent equilateral triangles centered at $x=0$ and having one vertex on the positive axis of reals.

entiating these the requisite number of times, obtain series converging uniformly to zero in and on $T_2$. Adding these series to (1) produces still other series of the type (1) converging uniformly to $f(x)$ in and on the smaller of $T_1$ and $T_2$ but having coefficients different from the formal coefficients.

It appears from the above that the expansion (1) is by no means unique, there being infinitely many expansions of this type, all converging uniformly to $f(x)$ in a common triangle, which, in case the singularities of $f(x)$ are all more distant from the origin than $\pi$, has one vertex as close to $\pi$ as we care to place it. It is of interest to note the difference between this situation and that which obtains with power series, in which the coefficients are unique however small the circle of convergence, and with Fourier series, in which the coefficients are unique if the series is made to converge to the given function over the interval $(0, 2\pi)$ except possibly at a finite number of points. Of course, if the expansion (1) converges uniformly to $f(x)$ in the interval $0 \leqq x \leqq \pi$, the coefficients are unique.

3. **Extension of triangle of uniform convergence.** Let us suppose that one at least of the power series for the $\phi$'s fails to converge for $x=a$ and that all three series converge for $|x|<|a|$. Laying aside for the moment the cases in which $|a| \geqq \pi$, let us suppose $|a|<\pi$. Construct the circle $|x|=|a|$, and let $x=b$ be the point where this circle cuts the axis of reals, $0<b<\pi$. Let $T_b$ be the equilateral triangle with one vertex at $x=b$. Then it is known that series (3) converge uniformly to the respective functions standing on the right-hand sides of equations (3) in and on the boundary of every closed region interior to $T_b$.* Let $T_a$ be the equilateral triangle with one side through $x=a$, and $T_\pi$ the equilateral triangle with one vertex at $\pi$.

THEOREM I. *The formal series* $\sum_{k=1}^{\infty} a_{ik}u_{ik}(x)$ *converges uniformly to* $x^{i-1}\phi_i(x^3)$ *in and on the boundary of every closed region interior to the smaller of* $T_a$ *and* $T_\pi$.

Let $z$ be that vertex of $T_a$ which is on the positive axis of reals if that vertex lies to the left of $\pi$, otherwise let $z=\pi$. Choose $y$ on this axis, $0<y<z$. We shall establish uniform convergence in the interval $0 \leqq x \leqq \alpha < y$, from which the theorem follows immediately.†

We examine the magnitudes of the coefficients $a_{ik}$. We have

$$\int_0^\pi x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx = \left[\int_0^y + \int_y^\pi\right] x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx.$$

But

* Ward, loc. cit., Theorem I, p. 742.

† Ward, loc. cit., Theorem II, p. 742.

$$\left| \int_y^\pi x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx \right| < 3M_1 \mid e^{\omega_3\rho_{ik}(\pi-y)} \mid ,$$

where $M_1 = \max |x^{i-1}\phi_i(x^3)|$ on the interval $0 \leqq x \leqq \pi$. Also

$$\begin{aligned}
&\int_0^y x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx \\
&= e^{\omega_1\rho_{ik}\pi}\int_0^y x^{i-1}\phi_i(x^3)\, e^{-\omega_1\rho_{ik}x}dx - \omega_2 e^{\omega_2\rho_{ik}\pi}\int_0^y x^{i-1}\phi_i(x^3)e^{-\omega_2\rho_{ik}x}dx \\
&\qquad - \omega_3 e^{\omega_3\rho_{ik}\pi}\int_0^y x^{i-1}\phi_i(x^3)\, e^{-\omega_3\rho_{ik}x}dx \\
&= e^{\omega_1\rho_{ik}\pi}\int_0^y x^{i-1}\phi_i(x^3)\, e^{\rho_{ik}x}dx + (-\omega_3)^{i+2}e^{\omega_2\rho_{ik}\pi}\int_0^{-\omega_2 y} x^{i-1}\phi_i(x^3)e^{\rho_{ik}x}dx \\
&\qquad + (-\omega_2)^{i+2}e^{\omega_3\rho_{ik}\pi}\int_0^{-\omega_3 y} x^{i-1}\phi_i(x^3)e^{\rho_{ik}x}dx \\
&= \delta_i(\rho_{ik}\pi)\int_0^{-y/2} x^{i-1}\phi_i(x^3)\, e^{\rho_{ik}x}dx + e^{\omega_1\rho_{ik}\pi}\int_{-y/2}^y x^{i-1}\phi_i(x^3)\, e^{\rho_{ik}x}dx \\
&\qquad + (-\omega_3)^{i+2}e^{\omega_1\rho_{ik}\pi}\int_{-y/2}^{-\omega_2 y} x^{i-1}\phi_i(x^3)\, e^{\rho_{ik}x}dx \\
&\qquad + (-\omega_2)^{i+2}e^{\omega_1\rho_{ik}\pi}\int_{-y/2}^{-\omega_3 y} x^{i-1}\phi_i(x^3)\, e^{\rho_{ik}x}dx.
\end{aligned}$$

These changes in the paths of integration, which are all taken as straight lines, can be made since the integrands are analytic functions of $x$ inside of $T_a$.

Now $\delta_i(\rho_{ik}\pi) = 0$.* The changes in the integral were made so that this fact could be used in showing that certain large portions of the integral cancel one another. Each of the three integrals whose lower limits are $-y/2$ is sufficiently small to enable us to draw the desired conclusions about the coefficients. In fact, at no point of the paths of integration does the exponential factor take on an absolute value greater than its absolute value at the upper limit. Hence

$$\left| \int_{-y/2}^{-\omega_i y} x^{i-1}\phi_i(x^3)e^{\rho_{ik}x}dx \right| < 3\pi M_2 \mid e^{-\omega_i\rho_{ik}y} \mid /2,$$

where $M_2 = \max |x^{i-1}\phi_i(x^3)|$ in and on the equilateral triangle centered at $x=0$ and having one vertex at the point $x=y$. Hence

* Ward, loc. cit., p. 741.

$$\left| \int_0^{\pi} x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx \right| < M\pi \mid e^{\omega_3\rho_{ik}(\pi-y)} \mid ,$$

where $M$ has an obvious constant value.

Since

$$\mid \delta_j(\rho_{ik}\pi) \mid > h \mid e^{\omega_3\rho_{ik}\pi} \mid \text{ if } i \neq j^*,$$

where $h$ is independent of $k$, we have

$$\mid a_{ik} \mid < M \mid e^{-\omega_3\rho_{ik}y} \mid /h.$$

If now $x$ is restricted to the interval $0 \leqq x \leqq \alpha < y$, we shall have

$$\begin{aligned} \mid a_{ik}\delta_i(\rho_{ik}x) \mid &< 3M \mid e^{\omega_3\rho_{ik}(x-y)} \mid /h \\ &\leqq 3M \mid e^{\omega_3\rho_{ik}(\alpha-y)} \mid /h \qquad (i = 1,2,3). \end{aligned}$$

This last expression, however, is the general term of a convergent series of positive constants. Consequently the series mentioned in the theorem converges to a function of $x$ analytic at every point interior to the smaller of $T_a$ and $T_\pi$. But it is known to converge to $x^{i-1}\phi_i(x^3)$ at every point interior to $T_b$. Hence, in view of the fact that $x^{i-1}\phi_i(x^3)$ is analytic at every interior point of $T_a$, the series converges uniformly to this function in and on the boundary of every closed region entirely within the smaller of $T_a$ and $T_\pi$.

If $|a| = \pi$, the triangle $T_b$ coincides with $T_\pi$, and Theorem I of part I of the 1927 paper tells us that $T_\pi$ is the triangle every closed interior region of which is a region of uniform convergence of the series.

We next consider uniform convergence of series (3) in the interval $0 \leqq x \leqq \pi$. For this it is evidently necessary that $\phi_i(\pi^3) = 0$. The following theorem gives sufficient conditions for this convergence.

THEOREM II. *If $x^{i-1}\phi_i(x^3)$ is analytic at every point inside $T_\pi$, continuous together with its first and second derivatives on $T_\pi$, and zero at $\pi$, then the formal series* (3) *converges uniformly to $x^{i-1}\phi_i(x^3)$ inside and on $T_\pi$.*

Since the method of proof is similar to that of Theorem I, we shall merely indicate the more important relations. Integrating by parts twice, we obtain

$$\int_0^{\pi} x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx = \int_0^{\pi} \psi_i(x)\delta_2[\rho_{ik}(\pi - x)]dx/\rho_{ik}^2,$$

where $\psi_i(x) = (d^2/dx^2)[x^{i-1}\phi_i(x^3)]$ is analytic at every interior point of $T_\pi$ and continuous on $T_\pi$. Also $\psi_i(-\omega_3 x) = (-\omega_3)^i\psi_i(x)$ and $\psi_i(-\omega_2 x) = (-\omega_2)^i\psi_i(x)$. Hence

* Ward, loc. cit., p. 719 (2).

$$\int_0^{\pi} x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx$$

$$= \frac{1}{\rho_{ik}^2}\Big[\delta_i(\rho_{ik}\pi)\int_0^{-\pi/2}\psi_i(x)\,e^{\rho_{ik}x}dx + e^{\omega_1\rho_{ik}\pi}\int_{-\pi/2}^{\pi}\psi_i(x)\,e^{\rho_{ik}x}dx$$

$$+ (-\omega_3)^{i+2}e^{\omega_1\rho_{ik}\pi}\int_{-\pi/2}^{-\omega_3\pi}\psi_i(x)e^{\rho_{ik}x}dx + (-\omega_2)^{i+2}e^{\omega_2\rho_{ik}\pi}\int_{-\pi/2}^{-\omega_3\pi}\psi_i(x)\,e^{\rho_{ik}x}dx\Big].$$

Consequently

$$\Big|\int_0^{\pi} x^{i-1}\phi_i(x^3)\delta_3[\rho_{ik}(\pi - x)]dx\Big| < M\pi/\rho_{ik}^2,\ |a_{ik}| < M\,|e^{-\omega_1\rho_{ik}\pi}|/(h\rho_{ik}^2),$$

and $|a_{ik}\delta_i(\rho_{ik}x)| < 3M/(h\rho_{ik}^2)$ if $x$ is within or on $T_\pi$. Since $1/\rho_{ik}^2$ is the general term of a convergent series of positive constants, this completes the proof.

Finally, we note that necessary conditions for the triangle of convergence to contain $\pi$ as an interior point are that $(d^{3n}/dx^{3n})[x^{i-1}\phi_i(x^3)]$ vanish at $x=\pi$ for $n=0, 1, 2, \cdots$. This fact and the above theorems enable us to state in most cases exactly the triangle of uniform convergence of the series (1) associated with a given analytic function. They fail to give us this information in case the parts $x^{i-1}\phi_i(x^3)$ of $f(x)$ have all their singularities at exterior points of $T_\pi$ and at the same time $(d^{3n}/dx^{3n})[x^{i-1}\phi_i(x^3)]$ vanish at $x=\pi$ for $n=0, 1, 2, \cdots$, and $i=1, 2, 3$; but we can say in such a case that the region of uniform convergence contains $T_\pi$.

4. **Differentiation and integration of formal series.** Inasmuch as the formal series (3) converge uniformly and their terms are analytic functions, differentiation term by term at any interior point of the region of uniform convergence is permissible, and the derived series will converge uniformly to the derived function. In this section we enquire whether if the series is differentiated term by term a number of times which is a multiple of three, the derived series will be the formal series of the derived function. It is clear in the case of series (4) that the third derived series are not the *formal* series of the third derived functions.

THEOREM III. *If $x^{i-1}\phi_i(x^3)$ is analytic in some closed region of which $x=0$ is an interior point, and has on the interval $0 \leqq x \leqq \pi$ a continuous fifth derivative, then a necessary and sufficient condition that the third derived series for this function be the formal series for the third derivative of the function is $\phi_i(\pi^3)=0$.*

The series for the given function being $\sum_{k=1}^{\infty} a_{ik}u_{ik}(x)$, the third derived

series is $-\sum_{k=1}^{\infty} a_{ik}\rho_{ik}^3 u_{ik}(x)$. Let us denote by $b_{ik}$ the $k$th coefficient in the formal series for the third derivative of the given function. Then

$$\pi b_{ik}\delta_{i-1}(\rho_{ik}\pi) = \int_0^{\pi} \frac{d^3}{dx^3}[x^{i-1}\phi_i(x^3)]\delta_3[\rho_{ik}(\pi - x)]dx.$$

The requirement that the given function have a continuous fifth derivative on the interval $(0, \pi)$ is made so that we may be sure that the formal series for the third derivative of this function converges to that third derivative.

Integrating by parts three times and making use of the formula for $a_{ik}$ gives

$$\pi b_{ik}\delta_{i-1}(\rho_{ik}\pi) = 3\rho_{ik}^2\pi^{i-1}\phi_i(\pi^3) - \pi\rho_{ik}^3 a_{ik}\delta_{i-1}(\rho_{ik}\pi),$$

from which the truth of the theorem is clear.

Since the formal series (3) converge uniformly in some region surrounding $x=0$, term-by-term integration along a simple curve lying within the region is admissible. We enquire whether, if the three-fold integral of $u_{ik}(x)$ is taken as $-u_{ik}(x)/\rho_{ik}^3$, the thrice integrated series will be the formal series of any function, and in case it is, of what function.

On integrating (3) we get

$$\int_0^x t^{i-1}\phi_i(t^3)dt = -\sum_{k=1}^{\infty} a_{ik}[\delta_{i+1}(\rho_{ik}x) - \delta_{i+1}(0)]/\rho_{ik}$$

or

$$\int_0^x t^{i-1}\phi_i(t^3)dt + C_{1i} = -\sum_{k=1}^{\infty} a_{ik}\delta_{i+1}(\rho_{ik}x)/\rho_{ik},$$

where $C_{1i} = -\delta_{i+1}(0)\sum_{k=1}^{\infty} a_{ik}/\rho_{ik}$. Two more integrations yield

$$\int_0^x\int_0^s\int_0^r t^{i-1}\phi_i(t^3)dtdrds + C_{1i}x^2/2 + C_{2i}x + C_{3i} = -\sum_{k=1}^{\infty} a_{ik}u_{ik}(x)/\rho_{ik}^3,$$

where $C_{2i} = \delta_{i+2}(0)\sum_{k=1}^{\infty} a_{ik}/\rho_{ik}^2$ and $C_{3i} = -\delta_i(0)\sum_{k=1}^{\infty} a_{ik}/\rho_{ik}^3$.

Now call

$$\phi(x) = \int_0^x\int_0^s\int_0^r t^{i-1}\phi_i(t^3)dtdrds$$

and expand $\phi(x)+C_{1i}x^2/2+C_{2i}x+C_{3i}$ in formal series. Since $C_{11}=C_{12}=C_{21}=C_{23}=C_{32}=C_{33}=0$, the formal expansion is of the form

$$\phi(x) + C_{1i}x^2/2 + C_{2i}x + C_{3i} = \sum_{k=1}^{\infty} b_{ik}u_{ik}(x), \tag{6}$$

where

$$\pi b_{ik}\delta_{i-1}(\rho_{ik}\pi) = \int_0^\pi [\phi(x) + C_{1i}x^2/2 + C_{2i}x + C_{3i}]\delta_3[\rho_{ik}(\pi - x)]dx,$$

and this series converges uniformly to the function which gave rise to it. Three integrations by parts give us

$$(7)\quad \pi b_{ik}\delta_{i-1}(\rho_{ik}\pi) = 3[\phi(\pi) + C_{1i}\pi^2/2 + C_{2i}\pi + C_{3i}]/\rho_{ik} - \pi a_{ik}\delta_{i-1}(\rho_{ik}\pi)/\rho_{ik}^3.$$

We shall now evaluate those $C$'s which are not given above as being zero. Placing $x=0$ and $i=1$ in (6) gives $C_{31}=3\sum_{k=1}^{\infty} b_{ik}$. But when $i=1$, (7) is

$$\pi b_{1k}\delta_3(\rho_{1k}\pi) = 3[\phi(\pi) + C_{31}]/\rho_{1k} - \pi a_{1k}\delta_3(\rho_{1k}\pi)/\rho_{1k}^3.$$

Hence, eliminating $b_{1k}$,

$$C_{31} = 3\sum_{k=1}^{\infty}\left[\frac{3\{\phi(\pi) + C_{31}\}}{\pi\rho_{1k}\delta_3(\rho_{1k}\pi)} - \frac{a_{1k}}{\rho_{1k}^3}\right]$$

$$= \frac{9}{\pi}\sum_{k=1}^{\infty}\frac{\phi(\pi)+C_{31}}{\rho_{1k}\delta_3(\rho_{1k}\pi)} + C_{31}.$$

Consequently

$$[\phi(\pi) + C_{31}]\sum_{k=1}^{\infty} 1/[\rho_{1k}\delta_3(\rho_{1k}\pi)] = 0.$$

Now the series $(3/\pi)\sum_{k=1}^{\infty}\delta_1(\rho_{1k}x)/[\rho_{1k}\delta_3(\rho_{1k}\pi)]$ converges uniformly to unity in every closed region within $T_\pi$. Hence, placing $x=0$, we have $\pi/9=\sum_{k=1}^{\infty}1/[\rho_{1k}\delta_3(\rho_{1k}\pi)]$, and therefore $C_{31}=-\phi(\pi)$.

Similarly, differentiating (6) with $i=2$ once and placing $x=0$ yields $C_{22}=-\phi(\pi)/\pi$, while differentiating twice with $i=3$ and placing $x=0$ yields $C_{13}=-2\phi(\pi)/\pi^2$. Hence, placing these results in (7), we have $b_{ik}=-a_{ik}/\rho_{ik}^3$. This we sum up in

THEOREM IV. *If the formal series for $x^{i-1}\phi_i(x^3)$ be integrated term by term three times with respect to $x$, the resulting series is the formal series for*

$$\phi(x) + C_{1i}x^2/2 + C_{2i}x + C_{3i},$$

*where*

$$\phi(x) = \int_0^x\int_0^s\int_0^r t^{i-1}\phi_i(t^3)dtdrds,\quad C_{11} = C_{12} = C_{21} = C_{23} = C_{32} = C_{33} = 0,$$

$$C_{13} = -2\phi(\pi)/\pi^2,\qquad C_{22} = -\phi(\pi)/\pi,\qquad C_{31} = -\phi(\pi).$$

5. **Relations involving the coefficients.** This section is concerned with

relations analogous to the Hurwitz theorem on the Fourier coefficients.* If the series $\sum_{k=1}^{\infty} a_{3k}u_{3k}(x)$ converges uniformly to $f(x)$ in $0 \leqq x \leqq \pi$, multiplying by $f(\pi-x)$ and integrating from 0 to $\pi$ yields

$$\pi \sum_{k=1}^{\infty} a_{3k}^2 \delta_2(\rho_{3k}\pi) = \int_0^{\pi} f(x)f(\pi - x)dx.$$

The series on the left hand side of this equation, however, will converge even if the formal series for $f(x)$ converges uniformly in an interval extending from 0 to slightly beyond $\pi/2$, and thus arises the question as to whether its sum is the same as when the interval of uniform convergence extends to $\pi$. This and similar questions are answered in the following theorem.

THEOREM V. *If the formal series* $\sum_{k=1}^{\infty} a_{ik}\delta_i(\rho_{ik}x)$ *converges uniformly to* $f(x)$ *in the interval* $0 \leqq x \leqq \pi/2+\delta$, *where* $\delta$ *is an arbitrary positive constant,*† *then*

(a) $$\pi \sum_{k=1}^{\infty} a_{1k}^2 \delta_3(\rho_{1k}\pi)/\rho_{1k}^2 = \int_0^{\pi}\int_{\pi}^{x}\int_{\pi}^{s} f(x)f(\pi - t)dtdsdx \quad \text{if } i = 1,$$

(b) $$\pi \sum_{k=1}^{\infty} a_{2k}^2 \delta_1(\rho_{2k}\pi)/\rho_{2k} = \int_0^{\pi}\int_{\pi}^{x} f(x)f(\pi - s)dsdx \quad \text{if } i = 2,$$

(c) $$\pi \sum_{k=1}^{\infty} a_{3k}^2 \delta_2(\rho_{3k}\pi) = \int_0^{\pi} f(x)f(\pi - x)dx \quad \text{if } i = 3.$$

The proofs of these are similar to one another, and for that reason we present the proof of (b) only. From our hypothesis it follows that $\sum_{k=1}^{\infty} a_{2k}\delta_2[\rho_{2k}(\pi-s)]$ converges uniformly to $f(\pi-s)$ in the interval $\pi/2 - \delta \leqq s \leqq \pi$. Hence

$$\begin{aligned}\int_{\pi/2}^{\pi}\int_{\pi}^{x} f(x)f(\pi - s)dsdx &= \int_{\pi/2}^{\pi} f(x) \sum_{k=1}^{\infty} a_{2k} \int_{\pi}^{x} \delta_2[\rho_{2k}(\pi - s)]dsdx \\ &= \int_{\pi/2}^{\pi}\left[f(x) \sum_{k=1}^{\infty} a_{2k}\delta_3[\rho_{2k}(x)]/\rho_{2k}\right]dx \\ &= \sum_{k=1}^{\infty} a_{2k} \int_{\pi/2}^{\pi} f(x)\delta_3[\rho_{2k}(\pi-x)]dx/\rho_{2k}.\end{aligned}$$

But, making use of (2) with $i=2$, this becomes

$$\int_{\pi/2}^{\pi}\int_{\pi}^{x} f(x)f(\pi - s)dsdx = \sum_{k=1}^{\infty} a_{2k}\left[\pi a_{2k}\delta_1(\rho_{2k}\pi) - \int_0^{\pi/2} f(x)\delta_3[\rho_{2k}(\pi - x)]dx\right]/\rho_{2k}$$

* Cf. Whittaker and Watson, *Modern Analysis*, second edition, §9.5.

† $\delta$ can be made zero if the term-by-term twice derived series converges uniformly to $f''(x)$ in $0 \leqq x \leqq \pi/2$.

$$= \pi \sum_{k=1}^{\infty} a_{2k}^2 \delta_1(\rho_{2k}\pi)/\rho_{2k} + \phi^2(\pi/2) + \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} \phi(x)\delta_2[\rho_{2k}(\pi - x)]dx,$$

where $\phi(x) = \int_0^x f(s)ds$. Also,

$$\int_0^{\pi/2} \int_{\pi}^{x} f(x)f(\pi - s)dsdx$$

$$= \int_0^{\pi/2} f(x)\left[\int_{\pi}^{\pi/2} f(\pi - s)ds + \int_{\pi/2}^{x} f(\pi - s)ds\right]dx$$

$$= -\phi^2(\pi/2) + \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} \delta_2(\rho_{2k}x) \int_{\pi/2}^{x} f(\pi - s)dsdx$$

$$= -\phi^2(\pi/2) + \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} \delta_3(\rho_{2k}x)f(\pi - x)dx/\rho_{2k}$$

$$= -\phi^2(\pi/2) + \sum_{k=1}^{\infty} a_{2k}\left[\pi a_{2k}\delta_1(\rho_{2k}\pi) - \int_0^{\pi/2} f(x)\delta_3[\rho_{2k}(\pi - x)]dx\right]/\rho_{2k}$$

$$= \pi \sum_{k=1}^{\infty} a_{2k}^2 \delta_1(\rho_{2k}\pi)/\rho_{2k} - \phi^2(\pi/2) - \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} f(x)\delta_3[\rho_{2k}(\pi - x)]dx/\rho_{2k}.$$

Hence

$$\int_0^{\pi} \int_{\pi}^{x} f(x)f(\pi - s)dsdx = 2\pi \sum_{k=1}^{\infty} a_{2k}^2 \delta_1(\rho_{2k}\pi)/\rho_{2k}$$

$$+ \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} \phi(x)\delta_2[\rho_{2k}(\pi - x)]dx \tag{8}$$

$$- \sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} f(x)\delta_3[\rho_{2k}(\pi - x)]dx/\rho_{2k}.$$

We are now in a position to insert the series developments for $f(x)$ and $\phi(x)$ since these converge uniformly in the interval over which we wish to integrate.

We have

$$\sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} \phi(x)\delta_2[\rho_{2k}(\pi - x)]dx = -\sum_{k=1}^{\infty} a_{2k} \sum_{l=1}^{\infty} a_{2l}c_{lk}/\rho_{2l},$$

where

$$c_{lk} = \int_0^{\pi/2} \delta_3(\rho_{2l}x)\delta_2[\rho_{2k}(\pi - x)]dx;$$

and

$$\sum_{k=1}^{\infty} a_{2k} \int_0^{\pi/2} f(x)\delta_3[\rho_{2k}(\pi - x)]dx/\rho_{2k} = \sum_{k=1}^{\infty} a_{2k} \sum_{l=1}^{\infty} a_{2l}d_{lk}/\rho_{2k},$$

where

$$d_{lk} = \int_0^{\pi/2} \delta_2(\rho_{2l}x)\delta_3[\rho_{2k}(\pi - x)]dx.$$

(9) Consider now the double series $\sum_{k,l=1}^{\infty} a_{2k}a_{2l}c_{lk}/\rho_{2l}$. Integration gives

$$c_{lk} = [\rho_{2k}^2\delta_3(\rho_{2l}\pi/2)\delta_3(\rho_{2k}\pi/2) + \rho_{2k}\rho_{2l}\delta_2(\rho_{2l}\pi/2)\delta_1(\rho_{2k}\pi/2) + \rho_{2l}^2\delta_1(\rho_{2l}\pi/2)\delta_2(\rho_{2k}\pi/2)]/(\rho_{2k}^3 - \rho_{2l}^3) \text{ if } l \neq k,$$

and $c_{kk} = \pi\delta_1(\rho_{2k}\pi)/2 - \delta_3^2(\rho_{2k}\pi/2)/\rho_{2k}$. Now

$$|\delta_j(\rho_{2l}\pi/2)| < 3e^{\rho_{2l}\pi/4}.$$

Hence

$$|c_{lk}| < 9e^{(\rho_{2l}+\rho_{2k})\pi/4}\frac{\rho_{2k}^2 + \rho_{2k}\rho_{2l} + \rho_{2l}^2}{|\rho_{2k}^3 - \rho_{2l}^3|}$$

$$= 9e^{(\rho_{2l}+\rho_{2k})\pi/4}/|\rho_{2k} - \rho_{2l}| \text{ if } l \neq k.$$

Also

$$|c_{kk}| < he^{\rho_{2k}\pi/2},$$

where $h$ is independent of $k$. Hence

$$|c_{lk}| < ge^{(\rho_{2l}+\rho_{2k})\pi/4}$$

where $k$ and $l$ are not necessarily different and $g$ is independent of $k$ and $l$.

Furthermore,

$$|a_{2k}| < me^{-\rho_{2k}x_1/2},$$

where $\pi/2 < x_1 < \pi/2 + \delta$ and $m$ and $x_1$ are independent of $k$.* Hence

$$|a_{2k}a_{2l}c_{lk}/\rho_{2l}| < m^2ge^{(\rho_{2l}+\rho_{2k})(\pi/4 - x_1/2)}.$$

This is recognized as the general term of a convergent double series. Hence series (9) converges absolutely.

* Ward, loc. cit., Lemma, p. 720.

In the same way the double series $\sum_{k,l=1}^{\infty} a_{2k}a_{2l}d_{lk}/\rho_{2k}$ can be shown to converge absolutely. We are justified, therefore, in summing series (9) first over $k$ and then over $l$ and adding to the series in the $d$'s. This results in

$$\sum_{k,l=1}^{\infty} a_{2k}a_{2l}(c_{kl} + d_{lk})/\rho_{2k}. \tag{10}$$

Now $c_{kl}+d_{lk}=0$ or $\pi\delta_1(\rho_{2k}\pi)$ according as $l$ is not or is equal to $k$, as is easily shown by noting that

$$c_{kl} + d_{lk} = \int_0^{\pi} \delta_3(\rho_{2l}x)\delta_2[\rho_{2k}(\pi - x)]dx \tag{11}$$

and performing the indicated integration.

Hence the double series (10) reduces to the simple series $\pi\sum_{k=1}^{\infty} a_{2k}^2\delta_1(\rho_{2k}\pi)/\rho_{2k}$, and this with (8) proves part (b).

We note that if the second derived series of the series of the theorem converges uniformly to $f''(x)$ in $0 \leqq x \leqq \pi$, a simpler relation can be derived. In fact, multiplying $f''(\pi-x)=\sum_{k=1}^{\infty} a_{2k}\rho_{2k}^2\delta_3[\rho_{2k}(\pi-x)]$ by $f(x)$ and integrating from 0 to $\pi$ gives

$$\int_0^{\pi} f(x)f''(\pi - x)dx = \pi \sum_{k=1}^{\infty} a_{2k}^2\rho_{2k}^2\delta_1(\rho_{2k}\pi).$$

However, this relation is not in general true if the interval of uniform convergence does not extend as far as $\pi$.

6. **Determination of formal coefficients.** The determination of the formal coefficients by rendering stationary certain integrals is of special interest in the expansion problems here considered because the integrals are similar to those of the preceding section.

THEOREM VI. *Let $f(x)$ be a function capable of development into series (3) for a definite value of $i$. If $S_{in}(x)=\sum_{k=1}^{n} b_{ik}\delta_i(\rho_{ik}x)$ and $R_{in}(x)=f(x)-S_{in}(x)$, where $n$ is regarded as a constant and the $b$'s as independent variables, then the following integrals have stationary values when the $b$'s are equal to the formal coefficients:*

(a) $$\int_0^{\pi}\int_{\pi}^{x}\int_{\pi}^{s} R_n(x)R_n(\pi - t)dtdsdx \quad \textit{in case } i = 1,$$

(b) $$\int_0^{\pi}\int_{\pi}^{x} R_n(x)R_n(\pi - s)dsdx \quad \textit{in case } i = 2,$$

(c) $$\int_0^{\pi} R_n(x)R_n(\pi - x)dx \quad \textit{in case } i = 3.$$

Again we shall give the proof of part (b) only, the proofs of the other parts being very similar to it. We have

$$\int_\pi^x R_n(\pi - s)ds = \int_\pi^x f(\pi - s)ds - \int_\pi^x \sum_{k=1}^n b_{2k}\delta_2[\rho_{2k}(\pi - s)]ds$$

$$= \int_\pi^x f(\pi - s)ds - T_n(x),$$

where $T_n(x) = \sum_{k=1}^n b_{2k}\delta_3[\rho_{2k}(\pi - x)]/\rho_{2k}$. Hence, denoting by $I_n$ the integral to be rendered stationary,

$$I_n = \int_0^\pi \int_\pi^x f(x)f(\pi - s)dsdx - \int_0^\pi f(x)T_n(x)dx$$

$$- \int_0^\pi \int_\pi^x S_n(x)f(\pi - s)dsdx + \int_0^\pi S_n(x)T_n(x)dx,$$

and

$$\frac{\partial I_n}{\partial b_{2k}} = - \int_0^\pi f(x)\frac{\partial T_n(x)}{\partial b_{2k}}dx - \int_0^\pi \int_\pi^x f(\pi - s)\frac{\partial S_n(x)}{\partial b_{2k}}dsdx$$

$$+ \int_0^\pi S_n(x)\frac{\partial T_n(x)}{\partial b_{2k}}dx + \int_0^\pi \frac{\partial S_n(x)}{\partial b_{2k}}T_n(x)dx.$$

But

$$\frac{\partial S_n(x)}{\partial b_{2k}} = \delta_2(\rho_{2k}x)$$

and

$$\frac{\partial T_n(x)}{\partial b_{2k}} = \delta_3[\rho_{2k}(\pi - x)]/\rho_{2k}.$$

Putting in these values and remembering (11) this becomes

$$\frac{\partial I_n}{\partial b_{2k}} = - \int_0^\pi f(x)\delta_3[\rho_{2k}(\pi - x)]dx/\rho_{2k} \tag{12}$$

$$- \int_0^\pi \int_\pi^x f(\pi - s)\delta_2(\rho_{2k}x)dsdx + 2\pi b_{2k}\delta_1(\rho_{2k}\pi)/\rho_{2k}.$$

But, by integration by parts,

$$\int_0^\pi \int_\pi^x f(\pi - s)\delta_2(\rho_{2k}x)dsdx = \int_0^\pi f(\pi - x)\delta_3(\rho_{2k}x)dx/\rho_{2k}.$$

Changing the variable of integration to $x'$, where $x' = \pi - x$, and substituting in (12) results in

$$\frac{\partial I_n}{\partial b_{2k}} = 2\left[ -\int_0^\pi f(x)\delta_3[\rho_{2k}(\pi - x)]dx + \pi b_{2k}\delta_1(\rho_{2k}\pi)\right]/\rho_{2k},$$

which, on comparison with (2), completes the proof.

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

# HYPERGEODESIC MAPPING OF A SURFACE ON A PLANE*

BY
ERNEST P. LANE

## 1. Introduction

The purpose of this paper is to study a projective analogue of the metric problem† of Beltrami, to map a surface of ordinary space on a plane so that the geodesics on the surface correspond to the straight lines in the plane. A two-parameter family of *hypergeodesics* on a surface in ordinary space is defined‡ by a differential equation of the same form as the equation of the geodesics, in which, however, the coefficients are arbitrary given functions of the parameters on the surface instead of being the well known formulas containing the Christoffel symbols that appear in the familiar equation of the geodesics. The problem under consideration, then, is to set up a one-to-one correspondence between the points of the surface and the points of a plane so that a given covariant family of hypergeodesics shall correspond to the straight lines of the plane.

In the metric theory it is customary not only to suppose that two corresponding points on the surface and in the plane have the same curvilinear coördinates, but also that these are the abscissa and ordinate of the point in the plane. In the projective theory here presented it is found to be convenient first to normalize the curvilinear coördinates on the surface according to Fubini, and then to suppose that the curvilinear coördinates in the plane are the same as on the surface, so that Wilczynski's theory of plane nets is employed in studying the correspondence.

In §2, the theory of Fubini's canonical form of the differential equations of a surface is briefly summarized in so far as it will be needed in this paper. In §3, the theory of hypergeodesics, including three examples, is similarly treated; and likewise in §4, Wilczynski's theory of plane nets.

The theory of hypergeodesic mapping is constructed in §5. Three examples of this mapping are discussed in §6. Green's congruentially asso-

* Presented to the Society, April 18, 1930; received by the editors January 25, 1930.

† Beltrami, Annali di Matematica, (1), vol. 7 (1865), pp. 185–204; or else, Opere Complete, vol. 1, pp. 262–280. See also Bianchi, *Lezioni*, 3d edition, vol. 1, p. 310. For other extensions of the problem of Beltrami, see Kasner, *The generalized Beltrami problem concerning geodesic representation*, these Transactions, vol. 4 (1903), pp. 149–152.

‡ Fubini and Čech, *Geometria Proiettiva Differenziale*, vol. 1, p. 133. Wilczynski, *Some generalizations of geodesics*, these Transactions, vol. 23 (1922), p. 283.

ciated net is connected with the theory in §7. Finally, another method of mapping a surface on a plane is suggested in §8.

## 2. Analytic definition of a surface

A surface in ordinary space can be defined, except possibly for a projective transformation, by a pair of linear homogeneous partial differential equations of the second order, whose coefficients satisfy certain integrability conditions. In this section we shall recapitulate as much as will be needed of the well known theory of the analytic representation of a surface by a pair of equations in Fubini's canonical form.

Let the parametric equations, in projective homogeneous coördinates, of a non-ruled surface $S$ in ordinary space, be

$$x^{(i)} = x^{(i)}(u,v) \qquad (i = 1, \cdots, 4).$$

If the asymptotic curves on the surface $S$ are parametric and if the proportionality factor of the homogeneous coördinates is suitably chosen, the coördinates $x$ satisfy a system of two equations in Fubini's canonical form,

$$\begin{aligned} x_{uu} &= px + \theta_u x_u + \beta x_v, \\ x_{vv} &= qx + \gamma x_u + \theta_v x_v \qquad (\theta = \log \beta\gamma). \end{aligned} \tag{1}$$

The coefficients of these equations satisfy three integrability conditions, namely,

$$\begin{aligned} \theta_{uvv} &= (\gamma\phi)_u + 2q_u + \theta_v\theta_{uv} - \beta\gamma\psi, \\ \theta_{uuv} &= (\beta\psi)_v + 2p_v + \theta_u\theta_{uv} - \beta\gamma\phi, \\ p_{vv} - \theta_v p_v + \beta q_v + 2q\beta_v &= q_{uu} - \theta_u q_u + \gamma p_u + 2p\gamma_u, \end{aligned} \tag{2}$$

where

$$\phi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v. \tag{3}$$

It is convenient to define $\pi, \chi$ by the formulas

$$\pi = p + \beta\psi, \qquad \chi = q + \gamma\phi. \tag{4}$$

The most general transformation which leaves Fubini's canonical form invariant is

$$x = c\bar{x}, \qquad \bar{u} = U(u), \qquad \bar{v} = V(v) \qquad (c = \text{const.}). \tag{5}$$

The effect of this transformation on system (1) is to produce another system of the same form whose coefficients, indicated by dashes, are given by the following formulas:

$$\begin{aligned} \bar{p} &= p/U'^2, \qquad \bar{\beta} = \beta V'/U'^2 \\ \bar{q} &= q/V'^2, \qquad \bar{\gamma} = \gamma U'/V'^2, \qquad \bar{\theta} = \theta - \log U'V'. \end{aligned} \tag{6}$$

## 3. Hypergeodesics

The aim of this section is to present that part of the general theory of hypergeodesics which will be essential for our purposes. Three examples of hypergeodesics will be given.

The two-parameter family of curves defined on an integral surface $S$ of system (1) by a differential equation of the form

$$v'' = A + 3Bv' + 3Cv'^2 + Dv'^3, \tag{7}$$

where accents indicate total differentiation with respect to $u$, and $A, B, C, D$ are given functions of $u, v$, are called *hypergeodesics*. These hypergeodesics are covariant to the surface, that is, are geometrically definable in terms of the surface, in case the coefficients in equation (7) are transformed by (5) according to the formulas

$$\begin{aligned} \overline{A} &= AV'/U'^2, & 3U'\overline{B} &= 3B - U''/U', \\ \overline{D} &= DU'/V'^2, & 3V'\overline{C} &= 3C + V''/V'. \end{aligned} \tag{8}$$

It will be observed in this connection that $A$, $D$ are cogredient with $\beta$, $\gamma$ respectively, while $3B$, $3C$ are cogredient with $\theta_u$, $\theta_v$.

Among the many examples of hypergeodesics that are known, there are three which we shall use. The first example is *the curves of a pencil of conjugate nets*, that is, a one-parameter family of conjugate nets every one of which has the property that at every point of the surface its tangents form with the tangents of a fundamental conjugate net the same cross ratio. The curvilinear differential equation of a pencil of conjugate nets can be written* in the form

$$dv^2 - \lambda^2 h^2 du^2 = 0, \tag{9}$$

where $\lambda$ is a function of $u, v$, and $h$ is a constant. Eliminating $h$ by differentiation, one obtains an equation of the form (7) with

$$A = D = 0, \qquad 3B = \lambda_u/\lambda, \qquad 3C = \lambda_v/\lambda. \tag{10}$$

In order that the pencil may be covariant to the surface it is necessary and sufficient that $\lambda$ be transformed by (5) according to the formula

$$\overline{\lambda} = \lambda V'/U'. \tag{11}$$

The second example of hypergeodesics is *the projective geodesics*, which are the extremals of the invariant integral, called *the projective arc length*,

$$\int_{u_1}^{u_2} (\beta\gamma v')^{1/2} du. \tag{12}$$

* Lane, *A general theory of conjugate nets*, these Transactions, vol. 23 (1922), p. 283.

For these curves one finds, on calculating Euler's equation, a differential equation of the form (7) with

$$A = D = 0, \qquad 3B = \theta_u, \qquad 3C = -\theta_v. \tag{13}$$

Finally, the third example is *the union curves of a congruence* $\Gamma_1$. Such a congruence consists of the lines joining each point $x$ of the surface $S$ to the corresponding point

$$-ax_u - bx_v + x_{uv},$$

where $a, b$ are functions of $u, v$. A union curve of this congruence is such that at each of its points its osculating plane contains the line of the congruence through the point. The differential equation of the union curves is of the form (7) with

$$A = -\beta, \qquad 3B = \theta_u - 2b, \qquad 3C = -\theta_v + 2a, \qquad D = \gamma. \tag{14}$$

In particular, the union curves of *the projective normal congruence* are obtained by placing $a=b=0$ in equations (14), and the union curves of *the first directrix congruence* of Wilczynski by placing therein $2a=\psi$, $2b=\phi$.

## 4. Plane nets

It is the purpose of this section to summarize the part that we shall need of Wilczynski's theory of plane nets. Then the curvilinear differential equation of the straight lines in the plane will be calculated in Wilczynski's notation, apparently for the first time.

If the three projective homogeneous coördinates $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ of a point $P_x$ in a plane are given as functions of two parameters $u, v$, these coördinates are solutions of a system of three equations of the form

$$\begin{aligned} x_{uu} &= ax_u + bx_v + cx, \\ x_{uv} &= a'x_u + b'x_v + c'x, \\ x_{vv} &= a''x_u + b''x_v + c''x, \end{aligned} \tag{15}$$

whose coefficients satisfy* the following six integrability conditions:

$$\begin{aligned} a_v + a''b &= a_u' + a'b' + c', \\ b_u'' + a''b &= b_v' + a'b' + c', \\ b_v + ab' + bb'' + c &= b_u' + a'b + b'^2, \\ a_u'' + a'b'' + aa'' + c'' &= a_v' + a''b' + a'^2, \\ c_v + ac' + bc'' &= c_u' + a'c + b'c', \\ c_u'' + b''c' + a''c &= c_v' + b'c'' + a'c'. \end{aligned} \tag{16}$$

* Wilczynski, *One-parameter families and nets of plane curves*, these Transactions, vol. 12 (1911), p. 474.

The effect of the transformation (5) when performed on system (15) is to produce another system of the same form whose coefficients, indicated by dashes, are given by the following formulas:

$$
\begin{aligned}
&U'\bar{a} = a - U''/U', \quad && \bar{b} = bV'/U'^2, && \bar{c} = c/U'^2,\\
(17)\qquad &\bar{a}' = a'/V', && \bar{b}' = b'/U', && \bar{c}' = c'/U'V',\\
&\bar{a}'' = a''U'/V'^2, && V'\bar{b}'' = b'' - V''/V', && \bar{c}'' = c''/V'^2.
\end{aligned}
$$

Moreover, the effect of the transformation

$$x = \lambda(u,v)\bar{x} \tag{18}$$

on system (15) is given by the formulas

$$
\begin{aligned}
&\bar{a} = a - 2\lambda_u/\lambda, && \bar{b} = b, && \lambda\bar{c} = -\lambda_{uu} + a\lambda_u + b\lambda_v + c\lambda,\\
(19)\qquad &\bar{a}' = a' - \lambda_v/\lambda, && \bar{b}' = b' - \lambda_u/\lambda, &&\\
& && \lambda\bar{c}' = -\lambda_{uv} + a'\lambda_u + b'\lambda_v + c'\lambda, &&\\
&\bar{a}'' = a'', && \bar{b}'' = b'' - 2\lambda_v/\lambda, &&\\
& && \lambda\bar{c}'' = -\lambda_{vv} + a''\lambda_u + b''\lambda_v + c''\lambda. &&
\end{aligned}
$$

The differential equation of the straight lines of the plane can be calculated by making use of the fact that $x$, $x'$, $x''$ must satisfy a linear relation when the accents indicate total differentiation with respect to $u$ along a straight line of the plane. We have at once

$$
\begin{aligned}
x' &= x_u + x_v v',\\
x'' &= x_{uu} + 2x_{uv}v' + x_{vv}v'^2 + x_v v''.
\end{aligned}
\tag{20}
$$

Substituting for the second derivatives in the second of these equations the expressions given for them in equations (15), we can express $x''$ as a linear combination of $x$, $x_u$, $x_v$. Regarding each of $x$, $x'$, $x''$ as so expressed, and equating to zero the third-order determinant of the coefficients of $x$, $x_u$, $x_v$ in their expressions, we obtain *the curvilinear differential equation of the straight lines in the plane*, namely,

$$v'' = -b + (a - 2b')v' - (b'' - 2a')v'^2 + a''v'^3. \tag{21}$$

## 5. Hypergeodesic mapping

In this section we solve the problem of mapping a surface, as defined in §2, on a plane so that the covariant hypergeodesics of §3 correspond to the straight lines in the plane. The coefficients of the differential equations of the plane net corresponding to the asymptotic curves on the surface are determined in terms of the coefficients in the equation of the hypergeodesics,

and their derivatives. It is found that these coefficients must satisfy two conditions in order that the mapping in question may be possible.

Let us suppose that we have a one-to-one correspondence between an integral surface $S$ of system (1) and an integral plane net $N$ of equations (15), or at least between certain regions of these, such that corresponding points have the same curvilinear coördinates $u, v$. Moreover, let us demand that a given family of covariant hypergeodesics (7) shall correspond to the straight lines (21). Conditions necessary and sufficient therefore are

$$b = -A, \qquad a - 2b' = 3B, \qquad b'' - 2a' = -3C, \qquad a'' = D. \tag{22}$$

It may be remarked that the left members of these equations are shown by the formulas (19) to be absolutely invariant under the transformation (18), as they should be. Moreover, these equations are shown by equations (8) and (17) to be absolutely invariant under the transformation (5).

*The net $N$ is the map of the asymptotic curves on the surface $S$. If $A=D=0$, the asymptotic curves on $S$ belong to the given family of hypergeodesics, and map into two one-parameter families of straight lines composing the net $N$.*

Let us attempt to calculate the coefficients of system (15) in terms of $A, B, C, D$ and the derivatives of these functions. To this end, let us solve equations (22) for $a, b, a'', b''$, and substitute the resulting expressions for these four coefficients in the integrability conditions (16), obtaining

$$\begin{aligned}
(2b' + 3B)_v - AD &= a_u' + a'b' + c', \\
(2a' - 3C)_u - AD &= b_v' + a'b' + c', \\
-A_v + b'(b' + 3B) - A(a' - 3C) + c &= b_u', \\
D_u + a'(a' - 3C) + D(b' + 3B) + c'' &= a_v', \\
c_v + c'(b' + 3B) - Ac'' &= c_u' + a'c, \\
c_u'' + c'(a' - 3C) + Dc &= c_v' + b'c''.
\end{aligned} \tag{23}$$

The first two of these equations imply

$$(b' + B)_v = (a' - C)_u. \tag{24}$$

Consequently, equations (19) show that if $\lambda$ is a solution of the equations

$$(\log \lambda)_u = b' + B, \qquad (\log \lambda)_v = a' - C, \tag{25}$$

then, dropping dashes from the transformed coefficients, we shall have

$$a' = C, \qquad b' = -B. \tag{26}$$

The first, third, and fourth of equations (23) now give expressions for $c'$, $c$, $c''$ respectively, while the last two of equations (23) give two conditions

that must be satisfied by $A, B, C, D$. We reach thus the following conclusion.

*The hypergeodesics* (7) *on an integral surface S of system* (1) *can be mapped into the straight lines of the plane if, and only if, the coefficients A, B, C, D in equation* (7) *satisfy the two conditions*

$$(27)\quad \begin{aligned} &2B_{uv} - A_{vv} - C_{uu} - 6BB_v + 3AC_v - A_vD - 2AD_u + 3BC_u + 3CA_v = 0, \\ &-2C_{uv} + D_{uu} + B_{vv} - 6CC_u + 3DB_u - D_vA - 2DA_v + 3CB_v + 3BD_u = 0. \end{aligned}$$

*The parametric asymptotic curves on the surface S map into an integral net N of the system* (15) *whose coefficients are given by the formulas*

$$(28)\quad \begin{aligned} &a = B, \quad b = -A, \quad c = A_v - B_u + 2B^2 - 2AC, \\ &a' = C, \quad b' = -B, \quad c' = B_v - C_u - AD - BC, \\ &a'' = D, \quad b'' = -C, \quad c'' = -D_u + C_v + 2C^2 - 2BD. \end{aligned}$$

*The integrability conditions of this system* (15) *are all satisfied.*

## 6. Examples of hypergeodesic mapping

Three examples of hypergeodesic mapping, corresponding to the three special kinds of hypergeodesics mentioned in §3, will now be discussed.

For the first example, let us consider the curves of a pencil of conjugate nets. The conditions (27) reduce, in virtue of equations (10), to

$$(29)\qquad \mu_{uuv} - \mu_u\mu_{uv} = 0, \qquad \mu_{uvv} - \mu_v\mu_{uv} = 0,$$

where $\mu$ is defined by placing

$$(30)\qquad \mu = \log \lambda.$$

Integrating each of equations (29), one obtains

$$(31)\qquad \mu_{uu} - \tfrac{1}{2}\mu_u^2 = U_1, \qquad \mu_{vv} - \tfrac{1}{2}\mu_v^2 = V_1,$$

where $U_1$, $V_1$ are arbitrary functions of $u$ alone and of $v$ alone, respectively. By means of equation (11) it is possible to show that a transformation (5) for which we have

$$(32)\quad \begin{aligned} \mu_{uu} - \tfrac{1}{2}\mu_u^2 &= U'''/U' - 3U''^2/2U'^2, \\ \mu_{vv} - \tfrac{1}{2}\mu_v^2 &= V'''/V' - 3V''^2/2V'^2, \end{aligned}$$

reduces each of $U_1$ and $V_1$ to zero. Incidentally, the most general transformation (5) which leaves these conditions undisturbed has $U$ and $V$ equal to proper linear fractional functions with constant coefficients so that

$$(33)\qquad U = (au + b)/(cu + d), \qquad V = (a'v + b')/(c'v + d').$$

We now have

$$(34)\qquad \mu_{uu} + \tfrac{1}{2}\mu_u^2 = 0, \qquad \mu_{vv} - \tfrac{1}{2}\mu_v^2 = 0.$$

Integration of these equations leads to the formula for $\lambda$

$$\lambda = 1/(kuv + lu + mv + n)^2, \tag{35}$$

in which $k, l, m, n$ are constants. Therefore, *the curves of a pencil of conjugate nets* (9) *with* $\lambda$ *given by the formula* (35) *map into the straight lines of the plane.*

Two remarks concerning this result may be permitted. A pencil of the kind described is a pencil of isothermally conjugate nets, characterized by the condition $(\log \lambda)_{uv}=0$, if, and only if, $k=0$. If $k\neq 0$, it is possible, by means of a suitably chosen transformation (33), to make $l=m=0$.

For the second example of hypergeodesic mapping, let us consider the projective geodesics. The conditions (27) reduce, in virtue of equations (13), to

$$\theta_{uuv} - \theta_u\theta_{uv} = 0, \qquad \theta_{uvv} - \theta_v\theta_{uv} = 0. \tag{36}$$

Proceeding step by step as in the preceding example, we find that *the projective geodesics on an integral surface of system* (1) *map into the straight lines of the plane if, and only if,*

$$\beta\gamma = 1/(kuv + lu + mv + n)^2, \tag{37}$$

*where* $k, l, m, n$ *are constants.* If $k\neq 0$, we can make $l=m=0$.

For the third example of hypergeodesic mapping, let us consider the union curves of the projective normal congruence. We shall not write here the result of substituting in the conditions (27) the general expressions given for $A, B, C, D$ in equations (14). However, when $a=b=0$, this substitution leads to

$$\begin{aligned} \theta_{uuv} - \theta_u\theta_{uv} + (\beta\psi)_v + \beta\gamma\phi = 0, \\ \theta_{uvv} - \theta_v\theta_{uv} + (\gamma\phi)_u + \beta\gamma\psi = 0, \end{aligned} \tag{38}$$

where $\phi, \psi$ are defined in equations (3). By means of the first two integrability conditions (2), one obtains at once

$$(p+\beta\psi)_v=0, \quad (q+\gamma\phi)_u=0 \tag{39}$$

and on integrating these equations one arrives at

$$p + \beta\psi = U_1, \qquad q + \gamma\phi = V_1. \tag{40}$$

By means of a suitably chosen transformation (5) it is possible to reduce each of $U_1$ and $V_1$ to unity, so that $\pi=\chi=1$, the functions $\pi, \chi$ being defined in equations (3). Therefore, *the union curves of the projective normal congruence of an integral surface of a system* (1) *for which*

$$p = 1 - \beta\psi, \qquad q = 1 - \gamma\phi, \tag{41}$$

*map into the straight lines of the plane.*

A consequence of this result may be deduced. The differential equation of *the projective lines of curvature*, which are the curves corresponding to the developables of the projective normal congruence, is

$$\pi du^2 - \chi dv^2 = 0. \tag{42}$$

Therefore, *on a surface of the kind under consideration here the projective lines of curvature form an isothermally conjugate net*, since $\pi = \chi = 1$.

## 7. The congruentially associated net

The congruentially associated net of a given plane net was defined* by Green as follows. As a point $P_x$ varies along a curve of the congruentially associated net, the harmonic conjugate of the tangent of this curve, with respect to the two tangents of the given net, passes through the corresponding focal point of the ray of the given net. The curvilinear differential equation of the congruentially associated net of an integral net of system (15) is

$$[c + b'(a - b') + a'b - b_u']du^2 + (a_u' - b_v')dudv - [c'' + a'(b'' - a') + a''b' - a_v']dv^2 = 0. \tag{43}$$

By means of equations (28) this equation reduces to

$$(A_v - 3AC)du^2 + (B_v + C_u)dudv + (D_u + 3BD)dv^2 = 0. \tag{44}$$

This is *the differential equation of the congruentially associated net of the net N into which the asymptotics on the surface S map when the hypergeodesics on S map into the straight lines in the plane of the net N.*

It is easy to prove the following theorems. *In the case of the curves of a pencil of conjugate nets the congruentially associated net coincides with the net N, provided that the nets of the pencil of conjugate nets are not isothermally conjugate. If they are isothermally conjugate, the congruentially associated net is indeterminate. In the case of the projective geodesics the congruentially associated net is indeterminate.*

When the union curves of a congruence $\Gamma_1$ are mapped into the straight lines in a plane, the congruentially associated net of the net $N$ has the differential equation

$$\beta(\psi - 2a)du^2 + 2(b_v - a_u)du\,dv/3 - \gamma(\phi - 2b)dv^2 = 0. \tag{45}$$

The coefficients of $du^2$ and $dv^2$ vanish if the congruence $\Gamma_1$ is the first directrix congruence of Wilczynski. In this case the conditions (27) reduce to

* Green, *Plane nets with equal invariants*, Annals of Mathematics, vol. 19 (1918), p. 248.

$$\phi_{uv} + \phi_v(\log \gamma)_u - 3\beta\gamma\phi = 0,$$
$$\psi_{uv} + \psi_u(\log \beta)_v - 3\beta\gamma\psi = 0.$$

We reach thus the following theorem.

*If the surface $S$ is not isothermally asymptotic, so that $\phi_v - \psi_u \neq 0$, the first directrix congruence of Wilczynski is the only congruence $\Gamma_1$ whose union curves map into the straight lines of a plane so that the congruentially associated net of the net $N$ corresponding to the asymptotics coincides with the net $N$. If the surface $S$ is isothermally asymptotic, the congruentially associated net of the net $N$ is indeterminate in the case of the union curves of the directrix congruence.*

Placing $a = b = 0$ in equation (45) to obtain the differential equation of the congruentially associated net in the case of the union curves of the projective normal congruence, one sees that, *in this case, provided $\phi\psi \neq 0$, at each point of the plane the tangents of the congruentially associated net of the net $N$, corresponding to the asymptotics on the surface $S$, separate harmonically the tangents of the net $N$.* The cases $\phi\psi = 0$ are easily disposed of.

## 8. Another method of mapping

It may be proposed to map a surface onto a plane so that the hypergeodesics on the surface correspond to any two-parameter family of curves defined in the plane by a differential equation of the same form as that of the hypergeodesics. So far as the writer is aware, the projective differential aspects of such families of curves in the plane have not hitherto been studied. It would seem, therefore, that the following example of such a family of curves covariant to a given plane net may be of interest.

Equations (17), (19) show that the differential form $a''bdudv$ is absolutely invariant under the most general transformation leaving invariant an integral net of system (15). Therefore the integral

$$\int_{u_0}^{u} (a''bv')^{1/2} du$$

has a geometric significance. Euler's equation for the extremals of this integral is of the same form as equation (7), with

$$(46) \qquad A = D = 0, \qquad 3B = (\log a''b)_u, \qquad 3C = -(\log a''b)_v.$$

*These extremals constitute a family of curves covariant to the integral net of system* (15).

The projective geodesics on an integral surface $S$ of system (1) map into these curves in case

$$(47) \qquad (\log \beta\gamma)_u = (\log a''b)_u, \qquad (\log \beta\gamma)_v = (\log a''b)_v.$$

Integrating, and absorbing a constant factor into the product $a''b$, we see that *the projective geodesics on the surface S map into the plane curves given by equation* (46) *and covariant to the plane net N corresponding to the asymptotic curves on S if, and only if,*

$$a''b' = \beta\gamma. \tag{48}$$

It may be remarked that if the plane net $N$ is perspective to the asymptotic net on the surface $S$, then the center of perspectivity and the plane of the net $N$ may be chosen to be the point (0, 0, 0, 1) and the plane $x_4=0$. Then it follows that $\beta=b$, $\gamma=a''$, so that *the perspectivity is a mapping of the kind under consideration here.*

It is clear that the bundle of planes with center at the center of perspectivity is a family of $\infty^2$ planes cutting the surface $S$ in $\infty^2$ plane curves which are union curves of the congruence $\Gamma_1$ which is the bundle of lines with its center at the center of perspectivity. Moreover, these planes cut the plane of the net $N$ in the straight lines of this plane. Therefore we have here *an instance of the mapping of the union curves of a congruence* $\Gamma_1$ *into the straight lines of the plane.* It may be shown, moreover, that if a family of hypergeodesics on the surface $S$ corresponds to the straight lines of the plane in the perspectivity aforesaid, then the hypergeodesics are the union curves of the bundle of lines with center at the center of perspectivity, and are precisely the plane curves just described on the surface $S$.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

# MANIFOLDS OF FUNCTIONS DEFINED BY SYSTEMS OF ALGEBRAIC DIFFERENTIAL EQUATIONS*

BY
J. F. RITT

This paper is concerned with the development of a theory of systems of algebraic ordinary differential equations, analogous to the theory of algebraic manifolds.

We deal with any finite or infinite system of algebraic differential equations in the independent variable $x$ and the dependent variables $y_1 \cdots y_n$. We write each equation in the form

$$F(x; y_1, \cdots, y_n) = 0,$$

where $F$ is a polynomial in the $y$'s and any number of their derivatives. The coefficients in $F$ will be supposed to be functions of $x$ meromorphic in some given open region $\mathfrak{A}$, and belonging to a given field $\mathfrak{F}$ of such functions. By a *field*, we understand a set of functions of $x$, not all zero, such that, given any function of the set, its derivative is also in the set, and such that, given any two functions, $f$ and $g$, of the set,

$$f + g,\ f - g,\ fg,\ \frac{f}{g} \quad (\text{when } g \neq 0)$$

are all in the set.†

An expression like $F$, above, will be called a *form*. With respect to every form introduced into our work, we shall assume, unless the contrary is stated, that its coefficients belong to $\mathfrak{F}$.

By a *solution* of a system of forms, we shall mean any set of functions, $y_1, \cdots, y_n$, analytic in some area contained in $\mathfrak{A}$, which cause all of the forms to vanish.‡ The totality of solutions of a system of forms will be called the *content* of the system. The content of any system will be called a *manifold*. If $\Sigma$ and $\Sigma'$ are two systems of forms such that every solution of $\Sigma$ is a solution of $\Sigma'$, then $\Sigma'$ will be said to *hold* $\Sigma$.

---

* Presented to the Society, April 18, 1930; received by the editors May 9, 1930.

† We do not assume, as Picard does in his construction of the Galois theory of linear differential equations, and as Landau does in his work on the factorization of linear differential forms, that $\mathfrak{F}$ contains all constants. See Picard, *Traité d'Analyse*, 2d edition, vol. 3, p. 562.

‡ An alternative definition of a solution would be any set of convergent series of powers of $x-a$, where $a$ is any point of $\mathfrak{A}$, which cause the forms to vanish.



BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

A system $\Sigma$ of forms will be called *irreducible* if, $G$ and $H$ being any two forms such that $GH$ holds $\Sigma$, either $G$ holds $\Sigma$ or $H$ holds $\Sigma$. A system which is not irreducible will be called *reducible*. The system of equations obtained by equating the forms of a system to zero, and also the manifold which is the content of the system of forms, will be called *reducible* or *irreducible* according as the system of forms is reducible or irreducible.*

We can now state the chief result of the first part of our paper. *Every manifold is composed of a finite number of irreducible manifolds.* That is, given any system of forms, $\Sigma$, there exist a finite number of *irreducible* systems, $\Sigma_1, \cdots, \Sigma_s$, such that $\Sigma$ holds every $\Sigma_i$, and that every solution of $\Sigma$ is a solution of some $\Sigma_i$. The decomposition into irreducible manifolds is essentially unique.

Let us consider an example. The equation

$$y'^2 - 4y = 0,$$

which has the general solution $y=(x-a)^2$ and a singular solution $y=0$, is a reducible system in the field of all constants. For neither $y''-2$ nor $y'$ vanishes for all solutions of the equation, while $(y''-2)y'$ does. The equation is equivalent to the two irreducible systems

$$y'^2 - 4y = 0, \quad y'' - 2 = 0,$$

and

$$y'^2 - 4y = 0, \qquad y' = 0.$$

The decomposition theorem follows from a lemma which bears a certain analogy to Hilbert's theorem on the existence of a finite basis for any infinite system of polynomials. We prove that *if $\Sigma$ is an infinite set of forms then $\Sigma$ has a finite subset whose content is identical with that of $\Sigma$.*†

In the second part of our paper, we investigate the structure of an irreducible manifold. We obtain, for every irreducible system, a differential equation which we call the *resolvent* of the system. Finding all non-singular

* The property that we have used in defining *irreducible manifold* is, of course, analogous to a characteristic property of irreducible algebraic manifolds. Of the different treatments of algebraic manifolds, that of van der Waerden, loc. cit., seems to be the only one that uses this property as a defining property. By the method of the present paper, the theorem that every algebraic manifold consists of a finite number of irreducible manifolds can be proved in a manner even simpler than that of van der Waerden, without using Lasker's theorem.

† This result is very different in nature from that of Tresse for systems of partial differential equations. See Tresse, Acta Mathematica, vol. 18 (1894), p. 4. Also, Drach, Annales de l'Ecole Normale, vol. 34 (1898), p. 292. In solving his system algebraically for certain derivatives, Tresse has necessarily to confine himself to a portion of the content of his system. The chief feature of the present investigation is its completeness from the algebraic standpoint.

solutions of the resolvent is equivalent to determining the content of the irreducible system.

To see what is contained in the idea of the resolvent, let us consider a differential equation $\alpha=0$, where $\alpha$ is a form in the $q+1$ indeterminates $w$; $u_1, \cdots, u_q$, irreducible, as a polynomial in $\mathfrak{F}$ in the indeterminates and their derivatives.

Let $\alpha$ be of order $r$ in $w$. Let $F=\partial\alpha/\partial w_r$, where $w_r$ is the $r$th derivative of $w$. Let $H$ be the coefficient of the highest power of $w_r$ in $\alpha$. We call a solution of $\alpha=0$, for which neither $F$ nor $H$ vanishes, a *regular solution*.

We prove that the totality of forms which vanish for all regular solutions of $\alpha=0$ is an irreducible system. The content of this system is one of the irreducible manifolds in the content of $\alpha$. We call this irreducible manifold the *general solution* of $\alpha$.

Now, suppose that we have $p$ rational combinations of $w; u_1, \cdots, u_q$ and their derivatives, with coefficients in $\mathfrak{F}$,

$$y_i = R_i(w; u_1, \cdots, u_q) \qquad (i = 1, \cdots, p), \tag{1}$$

no denominator vanishing for all regular solutions of $\alpha=0$.

We prove that there exist forms in $u_1, \cdots, u_q; y_1, \cdots, y_p$, which vanish for all $u$'s and $y$'s in (1), it being understood that $w; u_1, \cdots, u_q$ belongs to the general solution of $\alpha$. The totality of these forms in the $u$'s and $y$'s constitutes an irreducible system.

Conversely, let us consider any irreducible system in $y_1, \cdots, y_n$. A certain number, $q$, of the $y$'s are found to play the rôle of *arbitrary functions* in the content of the system. We call these $u_1, \cdots, u_q$, and designate the remaining $y$'s by $y_1, \cdots, y_p$ $(p+q=n)$.

We show that, if $\mathfrak{F}$ does not consist purely of constants, it is possible to form a rational combination $w$ of the $u$'s, $y$'s and their derivatives, in such a way that $y_1, \cdots, y_p$ become rational combinations of $w; u_1, \cdots, u_q$ and their derivatives. The new indeterminate, $w$, satisfies a differential equation

$$\alpha(u_1, \cdots, u_q; w) = 0,$$

with $\alpha$ irreducible as a polynomial in $\mathfrak{F}$. This equation is a *resolvent* of the irreducible system.

The introduction of the resolvent creates a perfect analogy between the notion of the content of a system of algebraic differential equations and the notion of algebraic function of several variables.

The resolvent can be used to advantage in the study of such questions as the influence on the reducibility of a system of the adjunction of new functions to $\mathfrak{F}$.

Certain features of our proofs indicate that a theory of ideals of differential forms underlies the manifold theory. We are at present investigating this question.*

One will notice that we do not furnish a method for resolving a system into irreducible systems, or a method for constructing the resolvent. These questions, also, we expect to treat in further communications.

Our work has, apparently, nothing in common with the researches of Riquier and others on the degree of generality of the solution of a system of partial differential equations.† We reserve for later papers the extension of our results to partial differential equations.‡

The irreducible differential equations of Koenigsberger,§ and Drach's irreducible systems of partial differential equations,¶ are irreducible in the sense explained above. The definitions of Koenigsberger and of Drach, which demand much more for irreducibility than does ours, are the starting points of group-theoretic investigations, which parallel the Galois theory.‖ Our definition leads, as we have seen, in a different direction.

This paper has a degree of contact with the work on field theory and elimination theory of the modern school of German algebraists. We would mention particularly the writings of Steinitz, Emmy Noether, Schmeidler and van der Waerden.**

## PART I. RESOLUTION OF A SYSTEM INTO IRREDUCIBLE SYSTEMS

### Classification of forms

1. Derivatives of functions $y_i$ will be indicated by means of a second subscript. Thus

$$y_{ij} = (d^j/dx^j)y_i.$$

We write, frequently, $y_i = y_{i0}$.

By the *jth derivative* of a form $F$, we mean the form obtained by differentiating $F$ $j$ times with respect to $x$, regarding $y_1, \cdots, y_n$ as functions of $x$.

By the *order of $F$ with respect to $y_i$*, if $F$ involves $y_i$ or some of its deriva-

* In this connection we have recently proved that if $G$ holds the system $F_1, \cdots, F_p$, then some power of $G$ is a linear combination of the $F$'s and their derivatives, with forms for coefficients. This is analogous to a theorem of Hilbert for polynomials. (Note added in proof, August 4, 1930.)

† See Janet, *Systèmes d'Equations aux Dérivées Partielles*, Paris, 1929.

‡ We have extended the theorem on the equivalence of a system to a finite number of irreducible systems to partial differential equations. (Note added in proof, August 4, 1930.)

§ *Lehrbuch der Differentialgleichungen*, Leipzig, 1889.

¶ Loc. cit., p. 295.

‖ In connection with Koenigsberger's definition, we have in mind the Picard-Vessiot theory.

** For references, see van der Waerden, Mathematische Annalen, vol. 97 (1927), p. 196.

tives effectively, we shall mean the greatest $j$ such that $y_{ij}$ is present in a term of $F$ with a coefficient distinct from zero. If $F$ does not contain $y_i$, the order of $F$ with respect to $y_i$ will be taken as zero.

By the *class* of a form which effectively involves some of the $y$'s, we shall mean the greatest $p$ such that some $y_{pj}$ is present in $F$. If $F$ is simply a function of $x$, $F$ will be said to be of class 0.

Let $F_1$ and $F_2$ be two forms. If $F_2$ is of higher order than $F_1$ in some $y_p$, $F_2$ will be said to be *of higher rank* than $F_1$ *in* $y_p$. If $F_1$ and $F_2$ are of the same order, say $q$, in $y_p$, and if $F_2$ is of greater degree than $F_1$ in $y_{pq}$,* then, again, $F_2$ will be said to be of higher rank than $F_1$ in $y_p$. Two forms for which no difference in rank is established by the foregoing criteria, will be said to be of the same rank in $y_p$.

If $F_2$ is of higher class than $F_1$, $F_2$ will be said to be *of higher rank* than $F_1$. If $F_2$ and $F_1$ are of the same class $p>0$, and if $F_2$ is of higher rank than $F_1$ in $y_p$, then, again, $F_2$ will be said to be of higher rank than $F$. Two forms for which no difference in rank is created by the preceding, will be said to be of the same rank.†

## Completeness of infinite systems

2. In §§2–11, we prove the following lemma:

Lemma. *Every infinite set of forms in* $y_1, \cdots, y_n$ *has a finite subset whose content is identical with that of the infinite set.*

An infinite system of forms whose content is identical with that of one of its finite subsets will be called *complete*.‡ Systems which are not complete will be called *incomplete*. In what follows, we assume the existence of incomplete systems, and force a contradiction.

3. We prove the following lemma:

Lemma. *Let* $\Sigma$ *be an incomplete system. Let a form* $F$, *not in* $\Sigma$, *and a subset* $\Sigma'$ *of* $\Sigma$ *exist, such that the system* $\Lambda$, *composed of the forms of* $\Sigma$ *not in* $\Sigma'$ *and of the products of the forms of* $\Sigma'$ *by* $F$, *is complete. Then the system* $\Sigma+F$, *obtained by adjoining* $F$ *to* $\Sigma$, *is incomplete.*

Suppose that $\Sigma+F$ is complete. Let

$$F;\quad G_1, \cdots, G_p; H_1, \cdots, H_q, \tag{2}$$

where the $H$'s, but not the $G$'s, belong to $\Sigma'$, be a subset of $\Sigma+F$ whose

* Considered as a polynomial in $y_{pq}$. If a form is identically zero (hence of order 0 in every $y_p$), it will be considered of degree 0 in every $y_{p0}$. This leads to no difficulties.

† Thus all forms of class 0 are of the same rank.

‡ If some finite subset has no solutions, the infinite set will be considered complete.

content is identical with that of $\Sigma+F$. We evidently may, and we shall, assume that the content of

$$G_1, \cdots, G_p;\ FH_1, \cdots, FH_q \tag{3}$$

is that of $\Lambda$. Now, let $K$ be a form of $\Sigma$ which does not hold

$$G_1, \cdots, G_p;\ H_1, \cdots, H_q. \tag{4}$$

As $FK$ holds (3), and as (3) holds (4), certain solutions of (4) which are not solutions of $K$ must be solutions of $F$. Thus $K$ does not hold (2). This proves the lemma.

4. By a *first form* of a system of forms, not all zero, we shall mean a form of the system, not zero, whose rank is not greater than that of any other non-zero form of the system.

From among all incomplete systems in $y_1, \cdots, y_n$, we select one whose first forms have a rank not greater than the rank of the first forms of any other incomplete system. Let $\Sigma_1$ be such an incomplete system, and let $\alpha_1$ be one of its first forms.

Let $\alpha_1$ be of class $p_1$. Then $p_1>0$, else $\alpha_1$ would have no solutions and $\Sigma_1$ would be complete.

5. Let a system $\Sigma$ contain a form $\alpha$ of class $p>0$. We call $\Sigma$ *reduced with respect to* $\alpha$ if every form of $\Sigma$, distinct from $\alpha$, is of lower rank than $\alpha$ in $y_p$.

6. We prove the following lemma:

LEMMA. *Given any incomplete system $\Sigma$ which contains the $\alpha_1$ of §4, there exists an incomplete system which has $\alpha_1$ for first form, is reduced with respect to $\alpha_1$, and contains every form of $\Sigma$ which is of lower rank than $\alpha_1$ in $y_{p_1}$.*

Let $\alpha_1$ be of order $r$ in $y_{p_1}$. The $q$th derivative of $\alpha_1$ will be of order $r+q$ in $y_{p_1}$ and will be linear in $y_{p_1,r+q}$, with $\partial\alpha_1/\partial y_{p_1,r}$ for coefficient of $y_{p_1,r+q}$. Let $F=\partial\alpha_1/\partial y_{p_1,r}$. Then $F$ is of lower rank than $\alpha_1$.

Now, $G$ being any form of $\Sigma$ of order higher than $r$ in $y_{p_1}$, it is possible, using the algorithm of division, to find a non-negative integer $m$, depending on $G$, such that, when we subtract from $F^mG$ a suitable linear combination of the derivatives of $\alpha_1$, with forms in $y_1, \cdots, y_n$ for coefficients, the remainder, call it $B$, is of order not greater than $r$ in $y_{p_1}$. Let such a $B$ be found for every $G$.

Let $\Omega$ be the system composed of all $B$'s and of all forms of $\Sigma$ whose order in $y_{p_1}$ does not exceed $r$. We are going to show that $\Omega$ is incomplete.

Let $\Lambda$ be the system composed of the forms $F^mG$ and the forms of $\Sigma$ whose order in $y_{p_1}$ does not exceed $r$.

Every $F^mG$ holds the system of two forms composed of its $B$ and $\alpha_1$. Every $B$ holds the system formed of its $F^mG$ and $\alpha_1$. Thus, if $\Omega$ were complete, $\Lambda$ would be complete. Now, if $m \geqq 1$, $F^mG$ and $FG$ hold each other. Thus, if $\Lambda$ were complete, the system obtained from $\Sigma$ by multiplying some subset of $\Sigma$ by $F$ would be complete. Then, by §3, $\Sigma+F$ would be incomplete. This is impossible, because $F$ is of lower rank than $\alpha_1$ and is not identically zero. Thus $\Omega$ is incomplete.

Let $H$ be the coefficient of the highest power of $y_{p_1,r}$ in $\alpha_1$. Then $H$ is of lower rank than $\alpha_1$.

Let $K$ be any form of $\Omega$ which is not zero and which is not of lower degree than $\alpha_1$ in $y_{p_1,r}$. An integer $m \geqq 0$ exists such that

$$H^mK = C\alpha_1 + D,$$

where $C$ and $D$ are forms in $y_1, \cdots, y_n$ and where $D$ is either zero or of lower degree than $\alpha_1$ in $y_{p_1,r}$. We see, as above, that the system composed of $\alpha_1$, the forms of $\Sigma$ which are of lower rank than $\alpha_1$ in $y_{p_1}$, and the $D$'s, is incomplete. Furthermore, this system is reduced with respect to $\alpha_1$. The lemma is proved.

7. Consider any $B$ of $\Omega$, and the $G$ from which it is obtained.

We say that, if $q > p_1$, then $B$ is not of higher rank than $G$ with respect to $y_q$.

For instance, let $\beta$ be a derivative of $\alpha_1$, of the same order as $G$ in $y_{p_1}$. Let

$$F^mG = A\beta + G_1,$$

where $G_1$ is of lower order than $G$ in $y_{p_1}$. Surely $G_1$ is not of higher order than $G$ in $y_q$. Suppose that $G$ and $G_1$ are of the same order, $h$, in $y_q$. If $G_1$ involved $y_{qh}$ in a higher power than $G$ does, then $A$ would contain the higher power of $y_{qh}$, so that $A\beta$ would contain the higher power of $y_{qh}$ multiplied by the derivative of highest order of $y_{p_1}$ in $\beta$. There would thus be terms in $A\beta$ which would not be balanced by the terms of $F^mG$ and $G_1$.

Similarly, consider any $D$ of the final system of §6, and the $K$ which corresponds to it. We see that, if $q > p_1$, $D$ will not be of higher rank than $K$ in $y_q$.

The observations of this section will be of great utility in §9.

8. Consider any incomplete system which has $\alpha_1$ for first form and is reduced with respect to $\alpha_1$. In such a system, there cannot be a non-zero form which is distinct from $\alpha_1$ and of class not exceeding $p_1$, for such a form would have to be of lower rank than $\alpha_1$.

It follows that $p_1 < n$.

Of the non-zero forms in the above system which are distinct from $\alpha_1$, there are some of lowest rank. Such forms, we call *second forms* of the system.

From among all incomplete systems which have $\alpha_1$ for first form, and are reduced with respect to $\alpha_1$, we choose one, $\Sigma_2$, whose second forms are of as low a rank as is possible. Let $\alpha_2$, of class $p_2 > p_1$, be a second form of $\Sigma_2$.

9. We prove the following lemma:

LEMMA. *Given any incomplete system $\Sigma$ which contains $\alpha_1$ and $\alpha_2$, there exists an incomplete system containing $\alpha_1$ and $\alpha_2$, which is reduced with respect to $\alpha_1$ and $\alpha_2$, and which contains all forms of $\Sigma$ which are of lower rank than $\alpha_2$ in $y_{p_2}$ and of lower rank than $\alpha_1$ in $y_{p_1}$.*

We note that the system whose existence is to be proved, being reduced with respect to $\alpha_1$, has $\alpha_2$ as a second form.

Let $\alpha_2$ be of order $r$ in $y_{p_2}$. Let $G$ be any form of order higher than $r$ in $y_{p_2}$. Let $F = \partial\alpha_2/\partial y_{p_2,r}$. Then there is an $m \geqq 0$ such that when a suitable linear combination of the derivatives of $\alpha_2$ is subtracted from $F^m G$, the remainder, $B$, has an order in $y_{p_2}$ not greater than that of $\alpha_2$. The system $\Omega$ composed of the $B$'s and the forms of $\Sigma$ not of higher order than $r$ in $y_{p_2}$ must be incomplete. If not, $\Sigma + F$ would be incomplete. Now $F$, like $\alpha_2$, is of lower rank than $\alpha_1$ in $y_{p_1}$. By §6, there would be an incomplete system with $\alpha_1$ for first form, reduced with respect to $\alpha_1$ and containing $F$. This is impossible if the class of $F$ does not exceed $p_1$, for then $F$ must be of lower rank than $\alpha_1$. It is impossible if the class of $F$ exceeds $p_1$, since $F$ is of lower rank than $\alpha_2$. Thus $\Omega$ must be incomplete.

Again, if $H$ is the coefficient of the highest power of $y_{p_2,r}$ in $\alpha_2$, we have, for any form $K$ of $\Omega$, distinct from zero and of degree in $y_{p_2,r}$ not less than that of $\alpha_2$,

$$H^m K = C\alpha_2 + D$$

with $D$ either zero or of lower degree than $\alpha_2$ in $y_{p_2,r}$.

We shall show that the $D$'s, $\alpha_2$ and the forms of $\Omega$ which are of lower rank than $\alpha_2$ in $y_{p_2}$ ($\alpha_1$ is among them) constitute an incomplete system $\Xi$.

For, if $\Xi$ were complete, $\Omega + H$ would be incomplete. By §6, there would exist a system containing $\alpha_1$, reduced with respect to $\alpha_1$, and containing $H$. As was seen above, this is impossible.

Proceeding now with $\alpha_1$ as in §6, and operating on the forms of $\Xi$ of rank in $y_p$ not less than that of $\alpha_1$, we obtain an incomplete system containing $\alpha_1$, $\alpha_2$ and all forms of $\Sigma$ which are of lower rank than $\alpha_2$ in $y_{p_2}$ and of lower rank than $\alpha_1$ in $y_{p_1}$, the system being reduced with respect to $\alpha_1$. Now this

system is also reduced with respect to $\alpha_2$, for, as was seen in §7, when we operate on a form of $\Xi$ with $\alpha_1$, the new form obtained has a rank with respect to $y_{p_2}$ not greater than that of the original form. This proves the lemma.

10. Evidently an incomplete system containing $\alpha_1$ and $\alpha_2$, and reduced with respect to $\alpha_1$ and $\alpha_2$, contains no non-zero form other than $\alpha_1$ and $\alpha_2$ whose class does not exceed $p_2$.

We conclude that $p_2 < n$.

In the incomplete systems of the type just described, we call those forms whose class exceeds $p_2$, and whose rank is as low as it can be, with this condition, *third forms*.

We select a system $\Sigma_3$ with a third form $\alpha_3$ of as low a rank as is possible. We operate as above, obtaining an incomplete system which contains $\alpha_1$, $\alpha_2$, $\alpha_3$ and is reduced with respect to $\alpha_1$, $\alpha_2$, $\alpha_3$. It follows, if $\alpha_3$ is of class $p_3$, that $p_3 < n$.

11. Continuing in this fashion, we find that there exists an infinite sequence of integers

$$p_1 < p_2 < p_3 < \cdots,$$

all less than $n$. This absurdity proves the truth of the fundamental lemma stated in §2.

## Non-existence of a Hilbert theorem

12. It might be conjectured that in every system $\Sigma$ there is a finite system of forms such that every form of $\Sigma$ is a linear combination of the forms of the finite system, and their derivatives, with forms for coefficients. We shall show that this is not so.

We consider forms in a single dependent variable, $y$, and represent the $n$th derivative of $y$ by $y_n$.

Consider the system

$$y_1y_2,\ y_2y_3,\ \cdots,\ y_ny_{n+1},\ \cdots.$$

We shall show that no form of this system with $n>1$ is linearly expressible in terms of the forms which precede it, and their derivatives.

We notice that all of the forms, and all of their derivatives, are homogeneous polynomials of the second degree in the $y$'s. Also if the weight of $y_iy_j$ is defined as $i+j$, the $p$th derivative of $y_iy_j$ will be isobaric, with its terms of weight $i+j+p$.

Now if

$$y_ny_{n+1} = A_1y_1y_2 + \cdots + A_{n-1}y_{n-1}y_n + B_1\frac{d}{dx}(y_1y_2) + \cdots,$$

with the $A$'s, $B$'s, etc., forms, the terms not independent of the $y$'s in the $A$'s, etc., may be cast out, for they produce terms of degree greater than 2.

Again, considering the weights of the various forms, we find that

$$y_n y_{n+1} = C_1 \frac{d^{2n-2}}{dx^{2n-2}}(y_1 y_2) + \cdots + C_{n-1}\frac{d^2}{dx^2}(y_{n-1}y_n) \tag{5}$$

with $C$'s independent of the $y$'s. Now the $(2n-2)$d derivative of $y_1y_2$ contains a term $y_1y_{2n}$, and none of the other derivatives in (5) yields such a term. We conclude that $C_1=0$. Continuing, we find every $C$ to be zero. This proves our statement.

## Irreducible systems

13. We prove the following fundamental theorem:

Theorem. *Given any system $\Sigma$ of forms in $y_1, \cdots, y_n$, there exist a finite number of irreducible systems, $\Sigma_1, \cdots, \Sigma_s$, such that $\Sigma$ holds every $\Sigma_i$, while every solution of $\Sigma$ is a solution of some $\Sigma_i$.*

Let the theorem be false for some system $\Sigma$. Then $\Sigma$ is reducible. Let $G_1$ and $G_2$ be two forms such that $G_1G_2$, but neither $G_1$ nor $G_2$, holds $\Sigma$. Now $\Sigma$ holds $\Sigma+G_1$, and $\Sigma+G_2$, and every solution of $\Sigma$, being a solution of $G_1$ or of $G_2$, is a solution of $\Sigma+G_1$ or of $\Sigma+G_2$.

Thus at least one of the systems $\Sigma+G_1$ or $\Sigma+G_2$ is reducible. If either of these systems is reducible, we call it a system of the *first class*. There must be a system of the first class which, when treated like $\Sigma$, yields either one or two reducible systems, obtained by adjoining two forms to $\Sigma$. The reducible systems obtained through two adjunctions, we call systems of the *second class*. Some of the systems of the second class, when treated like $\Sigma$, must yield reducible systems obtained from $\Sigma$ by three adjunctions. We call these systems of the *third class*. We proceed in this manner, forming systems of all classes.

There must be a system of the first class whose forms are contained in systems of all classes higher than the first. Let $\Sigma+H_1$, where $H_1$ is either $G_1$ or $G_2$, be such a system of the first class. One of the systems of the second class which contains the forms of $\Sigma+H_1$ must have its forms contained in systems of all classes higher than the second. Let $\Sigma+H_1+H_2$ be such a system. Let an $H_p$ be found in this way for every $p$. Then the system composed of

$$\Sigma, H_1, H_2, \cdots, H_p, \cdots$$

is incomplete. This proves our theorem. It will be noticed that the proof involves making an infinite number of selections.

### Uniqueness of resolution

14. We suppose, suppressing certain of the irreducible systems $\Sigma_i$, if necessary, that no $\Sigma_i$ holds a $\Sigma_j$ with $j \neq i$.

It is then possible to prove that *the systems $\Sigma_i$ are essentially unique*; that is, *if $\Omega_1, \cdots, \Omega_t$ is a second set of irreducible systems, none of which holds any other, each of which is held by $\Sigma$, and which are such that every solution of $\Sigma$ is a solution of some $\Omega_i$, then $s=t$, and every $\Omega_i$ holds, and is held by, some $\Sigma_i$.*

We shall show that there is some $\Omega_i$ which holds $\Sigma_1$. If there were not, then each $\Omega_i$ would have a form which would not hold $\Sigma_1$. Such forms being selected, their product would hold each $\Omega_i$, consequently $\Sigma$, thus $\Sigma_1$. This is impossible if $\Sigma_1$ is irreducible and none of the forms holds $\Sigma_1$.

Then let $\Omega_1$ hold $\Sigma_1$. Now $\Omega_1$, similarly, must be held by some $\Sigma_i$, which must be $\Sigma_1$, since no $\Sigma_i$ with $i \neq 1$ holds $\Sigma_1$. Thus $\Omega_1$ and $\Sigma_1$ hold each other. The uniqueness is proved.

## PART II. STRUCTURE OF AN IRREDUCIBLE MANIFOLD

### General solution of a single equation

15. We are going to study the content of a single form, $\alpha$, of class $n>0$. We assume that $\alpha$ is an *irreducible form*, that is, that $\alpha$ cannot be expressed as the product of two forms, each of class greater than 0, and each with coefficients in $\mathfrak{F}$.

It is our object to make precise the notion of the *general solution* of $\alpha=0$.

We write $y_n=y$, and, if $n>1$, we write $q=n-1$, $y_i=u_i$, $i=1, \cdots, n-1$.

Our definition of the general solution will appear, at first, to depend on the order in which the dependent variables in $\alpha$ are taken; at least, on the manner in which $y$ is selected from among the dependent variables. But it will turn out, finally, that the definition is actually independent of such order.

Let $\alpha$ be of order $r$ in $y$. Let $F=\partial\alpha/\partial y_r$, where $y_r$ is the $r$th derivative of $y$, and let $H$ be the coefficient of the highest power of $y_r$ in $\alpha$. A solution of $\alpha$ which is not a solution of $F$ or of $H$ will be called a *regular solution* of $\alpha$.

Let $A$ and $B$ be forms in $u_1, \cdots, u_q; y$, which are such that every regular solution of $\alpha$ is a solution of $AB$.

We shall prove that *either every regular solution of $\alpha$ is a solution of $A$ or every regular solution is a solution of $B$.*

From Part I (§6) we know that there exists a form $A_1$, of lower rank

than $\alpha$, such that every regular solution of $\alpha$ which satisfies either of the equations $A=0$ or $A_1=0$, satisfies the other. For $B$, similarly, we find a form $B_1$, of lower rank than $\alpha$.

If, then, we can show that either $A_1$ or $B_1$ is zero identically, our result will be proved.

Suppose that neither $A_1$ nor $B_1$ is identically zero. Consider any set of numerical values of $x$ and of $u_1, \cdots, u_q$; $y$ and their derivatives appearing in $\alpha$, $A_1$, $B_1$, for which $\alpha=0$. Let the value of $x$ be $a$. Suppose that neither $F$ nor $H$ vanishes for these numerical values.*

We construct functions $u_1, \cdots, u_q$ which have for themselves and for their derivatives, at $a$, the values indicated above. The existence theorem for differential equations assures us that $\alpha$ has a regular solution in which the $u$'s, $y$ and their derivatives have the indicated values at $a$.† This means that the set of numerical values makes either $A_1$ or $B_1$ vanish. All in all, $A_1B_1FH$ vanishes for all numerical values for which $\alpha$ vanishes. This means, since $\alpha$ is an irreducible form, that $A_1B_1FH$ is the product of $\alpha$ by a form. This cannot be, since none of $A_1$, $B_1$, $F$, $H$ can be divisible by $\alpha$. This completes the proof.

16. It follows immediately, from §15, that the system of all forms in $u_1, \cdots, u_q$; $y$, which vanish for all regular solutions of $\alpha$, is an irreducible system. The irreducible manifold which is the content of this irreducible system will be called the *general solution* of $\alpha=0$ (or of $\alpha$).

We show that *every solution of $\alpha$, for which $F$ does not vanish, belongs to the general solution.*

Let $A$ be any form which vanishes for all regular solutions. As above, let $\alpha$ be of order $r$ in $y$. Let $A_1$ be a form, not of order greater than $r$ in $y$, which differs from some $F^mA$ by a linear combination of the derivatives of $\alpha$. Some $H^nA_1$ equals the product of $\alpha$ by a form, plus a form $A_2$ of lower rank than $\alpha$. As $A_2$ vanishes for all regular solutions of $\alpha$, $A_2$, by an argument used in §15, must be identically zero. Thus $H^nA_1$ is divisible by $\alpha$. This means, since $\alpha$ is an irreducible form, and $H$ is not divisible by $\alpha$, that $A_1$ is divisible by $\alpha$. Thus $A_1$ holds $\alpha$. Hence $A$ vanishes for all solutions of $\alpha$ for which $F$ does not vanish. This proves our statement.

We shall prove that *the general solution of $\alpha$ is not contained in any other irreducible manifold of solutions of $\alpha$.*

Let $M$ be any irreducible manifold of solutions of $\alpha$ which contains the general solution. Those solutions in $M$ which are not in the general solution

* We are assuming that the coefficients in $\alpha$, $A_1$, $B_1$, $F$, $H$ are all analytic at $a$.

† That is, when the functions $u$ are constructed, we regard $\alpha=0$ as an equation in $y$.

make $F$ vanish. Let $B$ be any form which vanishes for every solution in the general solution. Then $BF$ vanishes for every solution in $M$. Since $F$ does not vanish for every solution in $M$, $B$ must. Thus $M$ is identical with the general solution.

We shall prove that *the definition of the general solution is independent of the order in which the indeterminates $y_1, \cdots, y_n$ are taken.*

Let $M_1, \cdots, M_s$ be $s>1$ irreducible manifolds, none containing any other, which make up the content of $\alpha$.* Suppose that, when $y=y_n$, the general solution is $M_1$, whereas, when $y=y_1$, the general solution is $M_s$.

Let $F'$ have, relative to $y_1$, the definition which $F$ has relative to $y_n$. Then every solution in $M_1$ must make $F'$ vanish. For, let $B$ be a form which vanishes for every solution in $M_s$, but not for every solution in $M_1$. Then, since $BF'$ vanishes for every solution of $\alpha$, $F'$ must vanish for every solution of $M_1$.

This means that every set of numerical values of $x$, the $y$'s and their derivatives, which makes $\alpha$ vanish, and which does not make $F$ vanish, makes $F'$ vanish; that is, for numerical values for which $\alpha$ vanishes, $FF'$ vanishes. For, according to the existence theorem for differential equations, a set of numerical values with $\alpha=0$ and $F\neq 0$ furnishes a solution in $M_1$.

Then, since $\alpha$ is an irreducible form, $FF'$ must be divisible, as a polynomial in the $y$'s and their derivatives, with coefficients in $\mathfrak{F}$, by $\alpha$. This is impossible, for neither $F$ nor $F'$ can be divisible by $\alpha$. Our statement is proved.

## Analytic constitution of the general solution†

17. A solution $\bar{u}_1, \cdots, \bar{u}_q; \bar{y}$ of $\alpha$, for which either $F$ or $H$ vanishes, will be said to be *semi-regular* if there exists a set of points, dense in the area, contained in $\mathfrak{A}$, in which the functions of the solution are analytic, such that, given any point $a$ of the set, any positive integer $m$, and any $\epsilon>0$, there exists a regular solution $u_1, \cdots, u_q; y$ (analytic at $a$) such that

$$(6)\quad |u_{ij}(a)-\bar{u}_{ij}(a)|<\epsilon,\ |y_j(a)-\bar{y}_j(a)|<\epsilon\ (i=1,\cdots,q;j=0,\cdots,m).$$

Here, $u_{ij}$ is the $j$th derivative of $u_i$, and $y_j$ the $j$th derivative of $y$ ($u_{i0}=u_i$, $y_0=y$).

*Any solution for which $H=0$, but for which $F$ does not vanish, is semi-regular.* This is an immediate consequence of the implicit function theorem (applied to $\alpha$ with respect to $y_r$) and of the theorem on the differentiability

* When $s=1$ we have our result immediately.

† The results of §§ 17, 18, and the analogous results of § 24, have contact with the remainder of the paper only in §28.

of the solution of a differential equation with respect to the constants of integration.

Let $A$ be any form in $u_1, \cdots, u_q; y$ with coefficients meromorphic in $\mathfrak{A}$. *The coefficients in $A$ need not belong to* $\mathfrak{F}$. Suppose that $A$ vanishes for every regular solution of $\alpha$. We shall prove that *$A$ vanishes for every semi-regular solution of $\alpha$.*

Consider any semi-regular solution $\bar{u}_1, \cdots, \bar{u}_q; \bar{y}$, and the points $a$ described above. Since the points are dense in an area, we can choose a point $a$ at which the coefficients in $A$ are analytic. Let this be done. When $\bar{u}_1, \cdots, \bar{u}_q; \bar{y}$ are substituted into $A$, $A$ becomes a function $\phi(x)$ of $x$, which is zero at $a$. This is because $A$ vanishes for all regular solutions, and because of the $m, \epsilon$ item in the definition of semi-regular solution. Again, $\phi'(x)$ must be zero at $a$, because the form obtained differentiating $A$ with respect to $x$ vanishes for every regular solution. Similarly, every derivative of $\phi(x)$ is zero at $a$. This proves that $A$ vanishes for the semi-regular solution.

If we restrict ourselves to forms $A$ with coefficients in $\mathfrak{F}$, we see that *the semi-regular solutions of $\alpha$ belong to the general solution.*

18. We are going to prove that *the general solution of $\alpha$ is composed of the regular solutions and of the semi-regular solutions.**

We denote by $\alpha_j$ the $j$th derivative of $\alpha$. If $\alpha$ is of order $r$ in $y$, then $\alpha_j$ is of order $r+j$ in $y$. It is linear in $y_{r+j}$, the coefficient of $y_{r+j}$ being $F$. Also the order of $\alpha_j$ in each $u$ effectively present in $\alpha$ exceeds the corresponding order of $\alpha$ by $j$.

We shall examine the system of equations

$$\alpha = 0, \ \alpha_1 = 0, \cdots, \alpha_s = 0, \tag{7}$$

where $s$ is any positive integer, considering the equations not as differential equations, but merely as algebraic equations among a set of indeterminates $u_{ij}, y_j$. That is, any set of functions $u_{ij}, y_j$, analytic in some area in $\mathfrak{A}$, and satisfying (7), will be considered as a solution of (7). We do not ask, for instance, that $y_j$ be the derivative of $y_{j-1}$.

We know from the theory of algebraic manifolds that the solutions of (7) form a finite number of irreducible manifolds. An irreducible manifold, here, is the totality of solutions of a set of algebraic equations in the $u_{ij}$'s and $y_j$'s appearing in (7) (coefficients in $\mathfrak{F}$) the set of equations being such that if $AB$ vanishes for all of its solutions, where $A$ and $B$ are polynomials

* W. Weltmann, Archiv der Mathematik und Physik, vol. 58(1876), p. 337, defined a solution of an equation of the first order as singular if it cannot be approximated by solutions distinct from itself.

with coefficients in $\mathfrak{F}$, then either $A$ vanishes for all solutions or $B$ does.* We may and shall assume that none of the irreducible manifolds contains any other.

One of these irreducible manifolds must contain the general solution of $\alpha$. That is, there is one irreducible manifold such that, $u_{10}, \cdots, u_{q0}; y_0$ being in the general solution of $\alpha$, the irreducible manifold contains a solution $u_{ij}, y_j$, with $u_{ij}$ the $j$th derivative of $u_{i0}$ and $y_j$ the $j$th derivative of $y_0$.

Suppose that this is not so. Let $M_1, \cdots, M_t$ be the irreducible manifolds of which the solutions of (7) are composed, and let $A_i$, $i=1, \cdots, t$, be a form which vanishes for every solution in $M_i$, but not for every regular solution of $\alpha$. As the general solution of $\alpha$ is an irreducible manifold, there are regular solutions which do not make $A_1 \cdots A_t$ vanish. This contradicts the fact that every solution of $\alpha$ gives a solution of (7).

We shall identify an irreducible manifold $M$ of the solutions of (7) which contains the general solution of $\alpha$.

We call any solution of (7) for which neither $F$ nor $H$ vanishes, a *regular* solution of (7).

The equations (7) define $y_r, \cdots, y_{r+s}$ in terms of $y, \cdots, y_{r-1}$ and the $u_{ij}$'s. That is, if we let $y, \cdots, y_{r-1}$ and the $u_{ij}$'s be any functions, analytic in an area in $\mathfrak{A}$, which do not render zero the resultant of $\alpha$ and $F$ with respect to $y_r$,† (7) determines $y_r, \cdots, y_{r+s}$, in succession, furnishing a regular solution.

It follows from the general theory of algebraic manifolds that there is only one irreducible manifold of solutions of (7) whose solutions do not all make $F$ vanish. This irreducible manifold, which contains the regular solutions of (7), is the manifold $M$ we have been seeking. Furthermore, in addition to the regular solutions of (7), $M$ contains those solutions of (7) which have the property that, in every area in which they are analytic, there is an area in which they can be approximated arbitrarily closely by a regular solution.‡

Suppose then that $\bar{u}_1, \cdots, \bar{u}_q; \bar{y}$ is a solution in the general solution of

---

* It should be emphasized that $A$ and $B$ involve only the indeterminates in (7), and not their derivatives.

† This resultant vanishes if either $F$ or $H$ vanishes.

‡ It is known that, given a system of algebraic functions $z_1, \cdots, z_t$, of several variables, the values of $z_1, \cdots, z_t$ where they are analytic, together with all sets of $t$ numbers which can be approximated arbitrarily closely by such values, form an irreducible manifold in the field of all complex numbers. A similar result holds when the coefficients in the equations which determine the $z$'s are not constants, but analytic functions, and when the values of the $z$'s are analytic functions.

$\alpha$. Consider any area $P$ in which the solution is analytic. Let an $m$ and an $\epsilon$ be assigned, as in the definition of semi-regular solution. Take $s>m$, and consider the corresponding system (7). Let a regular solution $\bar{u}_{ij}$, $\bar{y}_j$ of (7) be found, analytic in some area $P_1$ in $P$, such that, in $P_1$,

$$|\bar{u}_{ij}-\bar{u}_{ij}|<\epsilon,\ |\bar{y}_j-\bar{y}_j|<\epsilon \tag{8}$$

for all subscripts appearing in (7).* We may and shall suppose that $F$ and $H$ are distinct from zero throughout $P_1$, for the solution $\bar{u}_{ij}$, $\bar{y}_j$. Let $a$ be any point of $P_1$. Let functions $u_1, \cdots, u_q$ be taken, analytic at $a$, so that $u_{ij}(a)=\bar{u}_{ij}(a)$ for all subscript pairs appearing in (8). We notice that, for each $u_i$, $j$ assumes values at least as great as $m$.† Then, by (7), the differential equation $\alpha=0$ has a regular solution with $u_1, \cdots, u_q$ as just taken, and with $y_j(a)=\bar{y}_j(a)$, $j=0, \cdots, r+s$. Thus, for the given $m$ and $\epsilon$, any point in $P_1$ will serve as the point $a$ in the definition of semi-regular solution. Now, using $2m$ and $\epsilon/2$, we can find an area $P_2$, interior to $P_1$, any point $a$ of which can be used as above. Similarly, using $\epsilon/4$ and $4m$, we find an area $P_3$ in $P_2$, etc. There is a point $a$ which is interior to every $P_i$. Given any $\epsilon>0$, and any $m$, the differential equation $\alpha=0$ has a regular solution, analytic at $a$, for which (6) holds.

Thus, every solution of the general solution is either regular or semi-regular.

## THE BASIC EQUATIONS

19. We consider a system $\Sigma$ of forms in $y_1, \cdots, y_n$, not all zero. We assume that $\Sigma$ has solutions and that $\Sigma$ contains every form which holds $\Sigma$. For the rest, $\Sigma$ may be reducible or irreducible.

There may be some $y$, say $y_i$, such that no form of $\Sigma$ involves only $y_i$; that is, every form in which $y_i$ appears effectively also involves effectively some $y_j$ with $j\neq i$. If there exist such indeterminates $y_i$, let us pick one of them, arbitrarily, and call it $u_1$.

There may be a $y$, distinct from $u_1$, such that no form involves only $u_1$ and the new $y$. Let any such second $y$, if one or more exist, be denoted by $u_2$.

Continuing in this way, we find a set, $u_1, \cdots, u_q (q<n)$, such that no form of $\Sigma$ involves any of the $u$'s alone. Let the remaining indeterminates be represented now by $y_1, \cdots, y_p$, $p+q=n$. Then, given any $y_i$ among $y_1, \cdots, y_p$, there is a form in $\Sigma$ which involves only $y_i$ and the $u$'s.

It will be seen, in §26, that, when $\Sigma$ is irreducible, $q$ does not depend on the particular way in which the $u$'s may be selected.

* If a $u_{ik}$ appears effectively in $\alpha$, every $u_{ij}$ with $j\leqq k$ is regarded as present in $\alpha$. Similarly for $y$.

† No difficulty arises here if a $u_i$ is not present effectively in $\alpha$.

In what follows, we shall speak, generally, as if $u$'s exist. It will be easy to see, in every case, what slight modifications of language are necessary when there are no $u$'s.

Of all forms, not zero, in $\Sigma$, which involve no indeterminates other than $y_1$ and the $u$'s, let $\alpha_1$ have a minimum rank in $y_1$.

There exist forms (not zero), involving only $y_1$, $y_2$ and the $u$'s, which are of lower rank in $y_1$ than $\alpha_1$. For instance, any form involving only $y_2$ and the $u$'s is of this type. Of all such forms, let $\alpha_2$ have a minimum rank in $y_2$.

Continuing in this way, we find a sequence of forms,

$$\alpha_1, \alpha_2, \cdots, \alpha_p, \tag{9}$$

where

(I) $\alpha_i$ involves only the $u$'s and $y_1, \cdots, y_i$;

(II) $\alpha_1$ is of a minimum rank in $y_1$;

(III) for $i>1$, $\alpha_i$ is of lower rank in $y_j$ than $\alpha_j$, $j=1, \cdots, i-1$;

(IV) for $i>1$, $\alpha_i$ is not of greater rank in $y_i$ than any other form with the properties (I) and (III).

We shall call (9) a *basic system*.

If $\alpha_i$ is of order $r_i$ in $y_i$, we let $F_i=\partial\alpha_i/\partial y_{i,r_i}$. We designate the coefficient of the highest power of $y_{i,r_i}$ in $\alpha_i$ by $H_i$.

No $F_i$ can belong to $\Sigma$, for $F_i$ is of lower rank than $\alpha_i$ in $y_i$, of lower rank than $\alpha_{i-1}$ in $y_{i-1}$, etc. Similarly, no $H_i$ can belong to $\Sigma$.

A solution of the system (9) for which no $F_i$ or $H_i$ vanishes will be called a *regular solution* of (9).

We are going to show that *every regular solution of* (9) *is a solution of* $\Sigma$.

Consider any form $\beta$ of $\Sigma$ which involves only $y_1$ and the $u$'s. There exist an $m$ and an $s$ such that, when a suitable linear combination of $\alpha_1$ and its derivatives is subtracted from $F_1^mH_1^s\beta$, the remainder, call it $\gamma$, is of lower rank than $\alpha_1$ in $y_1$. Then $\gamma$, which belongs to $\Sigma$, must be identically zero. Hence every solution of $\alpha_1$ for which $F_1H_1$ does not vanish is a solution of $\beta$.

Consider any form $\beta$, of $\Sigma$, which involves only $y_1$, $y_2$ and the $u$'s. We find, as above, a $\gamma$, belonging to $\Sigma$, involving no $y_i$ with $i>2$ and of lower rank than $\alpha_2$ in $y_2$, such that every solution of $\alpha_2=0$, $\gamma=0$, for which $F_2H_2$ does not vanish is a solution of $\beta$. Furthermore, for $m$ and $s$ appropriate, $F_1^mH_1^s\gamma$ is a linear combination of $\alpha_1$ and its derivatives, plus a form $\delta$ of lower rank than $\alpha_2$ in $y_2$ and of lower rank than $\alpha_1$ in $y_1$. (See §7.) Then $\delta=0$. Thus every $u_1, \cdots, u_q$; $y_1$ for which $\alpha_1=0$, and for which $F_1H_1$ does not vanish, makes $\gamma=0$ for any $y_2$. Thus a solution of $\alpha_1=0$, $\alpha_2=0$, for which none of $F_1$, $F_2$, $H_1$, $H_2$ vanishes is a solution of $\beta$.

Continuing in this way, we see that every regular solution of (9) is a solution of $\Sigma$.

20. Suppose now that $\Sigma$ *is irreducible.* As no $F_i$ or $H_i$ holds $\Sigma$, the product of all $F$'s and $H$'s does not hold $\Sigma$.

It follows that (9) *has regular solutions.*

Furthermore, *if a form vanishes for all regular solutions of* (9), *the form is in* $\Sigma$. For, if $G$ is such a form,

$$GF_1 \cdots F_pH_1 \cdots H_p$$

holds $\Sigma$, so that $G$ holds $\Sigma$.

## The resolvent

21. From now on, we shall understand, unless the contrary is stated, that $\mathfrak{F}$ *contains at least one function which is not a constant.*

Let $\Sigma$ be reducible or irreducible, but not without solutions. We understand, as above, that $\Sigma$ contains every form which holds $\Sigma$.

We are going to show *the existence, in* $\mathfrak{F}$, *of functions*

$$\mu_1, \cdots, \mu_p$$

*and the existence of a form G, in the u's alone, such that, given two solutions of* $\Sigma$ *with the same u's,*

$$u_1, \cdots, u_q;\ y_1', \cdots, y_p',$$
$$u_1, \cdots, u_q;\ y_1'', \cdots, y_p'',$$

*for the u's of which G does not vanish, and in which, for some i,* $y_i'$ *is not identical with* $y_i''$, *then*

$$\mu_1(y_1' - y_1'') + \cdots + \mu_p(y_p' - y_p'')$$

*is not zero.**

We consider the system of forms obtained from $\Sigma$ by replacing each $y_i$ by a new indeterminate $z_i$. We take the system $\Omega$ composed of the forms of $\Sigma$, the forms in the $z$'s just described, and also the form

$$\lambda_1(y_1 - z_1) + \cdots + \lambda_p(y_p - z_p),$$

in which the $\lambda$'s are indeterminates. That is, $\Omega$ involves $3p+q$ indeterminates, namely, the $u$'s, $y$'s, $z$'s, $\lambda$'s.

Let $\Lambda$ be any irreducible system which $\Omega$ holds. We understand that $\Lambda$ contains every form which holds $\Lambda$.

Suppose that some one of the forms $y_i - z_i$ does not hold $\Lambda$. We shall

* Naturally, we assume that the two solutions have a common domain of analyticity.

prove that $\Lambda$ contains a non-zero form which involves no indeterminates other than the $u$'s and the $\lambda$'s.

Since $\Lambda$ has all forms of $\Sigma$, $\Lambda$ has, for $i=1, \cdots, p$, a form $\beta_i$ involving only $y_i$ and the $u$'s. Let $\beta_i$ be taken so as to be of a minimum rank in $y_i$. Let $\beta_i$ be of order $r_i$ in $y_i$ and put $F_i=\partial\beta_i/\partial y_{i,r_i}$. Similarly, let $\gamma_i$, $i=1, \cdots, p$, be a form of $\Lambda$, in $z_i$ and the $u$'s alone, which is of a minimum rank in $z_i$. Each $\gamma_i$ being of order $s_i$ in $z_i$, let $K_i=\partial\gamma_i/\partial z_{i,s_i}$.

Then no $F_i$ or $K_i$ is in $\Lambda$.

To fix our ideas, let us suppose that $y_1-z_1$ is not in $\Lambda$. Consider any solution of $\Lambda$ for which

$$(y_1 - z_1)F_1 \cdots F_pK_1 \cdots K_p$$

(which is not in $\Lambda$) does not vanish.

For such a solution, we have

$$\lambda_1 = \frac{\lambda_2(y_2 - z_2) + \cdots + \lambda_p(y_p - z_p)}{y_1 - z_1}. \tag{10}$$

From (10) we find, for the $j$th derivative of $\lambda_1$, an expression

$$\lambda_{1j} = R_j(\lambda_2, \cdots, \lambda_p; y_1, \cdots, y_p; z_1, \cdots, z_p), \tag{11}$$

in which $R_j$ is rational in $\lambda_2, \cdots, \lambda_p$, the $y$'s, $z$'s and the derivatives of the foregoing functions, with coefficients in $\mathfrak{F}$. The denominator in each $R_j$ is a power of $y_1-z_1$.

If an $R_j$ involves a derivative of $y_i$ of order higher than $r_i$, we can get rid of that derivative by using its expression in the derivatives of $y_i$ of order $r_i$ or less found from $\beta_i=0$. Similarly, we transform each $R_j$ so as to be of order not exceeding $s_i$ in $z_i$.

The new expression of each $R_j$, which will involve the $u$'s, will have a denominator which is a product of powers of $y_1-z_1$, $F_i$, $K_i$, $i=1, \cdots, p$.

Then, in (10) and (11), only a finite number of functions $y_{ik}$, $z_{ik}$ will appear. If we use a sufficiently large number of equations (11), we can, using rigorous principles of elimination, obtain from them an algebraic relation among the functions $\lambda_{ik}$, $u_{ik}$, with coefficients in $\mathfrak{F}$, which holds for any solution of $\Lambda$ which does not cause $y_1-z_1$, any $F_i$, or any $K_i$, to vanish. Let

$$D = 0$$

be such a relation, where $D$ is a form in the $u$'s and $\lambda$'s, with coefficients in $\mathfrak{F}$. Then

$$DF_1 \cdots F_pK_1 \cdots K_p(y_1 - z_1)$$

holds $\Lambda$, so that $D$ is in $\Lambda$. We have thus proved that $\Lambda$ has a form involving only the $u$'s and $\lambda$'s.

Let $\Lambda_1, \cdots, \Lambda_r$ be a set of irreducible systems such that $\Omega$ holds each of them and that every solution of $\Omega$ is a solution of one of them. We suppose each $\Lambda_i$ to contain every form which holds that $\Lambda_i$. Let $\Lambda_1, \cdots, \Lambda_s$ each not contain some of the forms $y_i - z_i$ and let $\Lambda_{s+1}, \cdots, \Lambda_r$ each contain all of the forms $y_i - z_i$. Let $D_i$ be a form in $\Lambda_i$, $i = 1, \cdots, s$, involving only the $u$'s and $\lambda$'s.

We wish to show the existence in $\mathfrak{F}$ of $p$ functions, $\mu_1, \cdots, \mu_p$, such that, when each $\lambda_i$ is replaced by $\mu_i$ in $E = D_1 \cdots D_s$, then $E$ does not vanish identically in the $u$'s.

Let $E$ be written as a polynomial in the $u$'s and their derivatives, with forms in the $\lambda$'s as coefficients. Let $K$ be one of the coefficients in $E$. If we can fix each $\lambda_i$ in $\mathfrak{F}$ so that $K$ does not vanish, our result will be established.

Consider any non-constant function $\zeta$ in $\mathfrak{F}$. Let $P$ be a circle in $\mathfrak{A}$ in which $\zeta$ is analytic and assumes no value more than once. Any function analytic in $P$ can be approximated arbitrarily closely, in any area interior to $P$, by a polynomial in $\zeta$, hence by a polynomial in $\zeta$ with rational coefficients. All polynomials in $\zeta$ with rational coefficients are in $\mathfrak{F}$.

Thus, if $K$ vanishes for all $\lambda$'s in $\mathfrak{F}$, $K$ vanishes if the $\lambda$'s are any functions analytic in $P$. This is certainly impossible. Thus, the required $\mu$'s exist.

The solutions of $\Omega$, for $\lambda_j = \mu_j$, $j = 1, \cdots, p$, will be solutions of the systems $\Lambda_i$ for $\lambda_j = \mu_j$. Now the solutions with $\lambda_j = \mu_j$ of each $\Lambda_i$, $i = 1, \cdots, s$, have $u$'s which cause to vanish the form $G$, obtained from $E$ by putting $\lambda_j = \mu_j$. The solutions of $\Lambda_{s+1}, \cdots, \Lambda_r$, even with $\lambda_j = \mu_j$, have $y_i = z_i$, $i = 1, \cdots, p$.

We have thus the result stated at the head of this §21.

22. From this point on, to the end of our paper, *we assume $\Sigma$ irreducible.*

Let $A$, $B$, $G$ be forms in the $u$'s and $y$'s, not in $\Sigma$, $G$ involving only the $u$'s, which are such that for any two distinct solutions of $\Sigma$, with the same $u$'s, for which neither $G$ nor $B$ vanishes, $A/B$ gives two distinct functions of $x$.

We have seen that, when $\mathfrak{F}$ does not consist entirely of constants, forms $A$, $B$, $G$ exist, that, in fact, one may take $B = 1$ and take $A$ free of the $u$'s. On the other hand, when $\mathfrak{F}$ contains only constants, there may be no $A$, $B$, $G$. Consider, for instance, the system

$$\frac{dy_1}{dx} = 0, \quad \frac{dy_2}{dx} = 0.$$

Any rational combination of $y_1$ and $y_2$ (and of their derivatives) with con-

stant coefficients, will have a single value for an infinite number of choices of $y_1$ and $y_2$.

We introduce a new indeterminate, $w$, and form a system $\Lambda$ by adjoining $Bw-A$ to $\Sigma$. Let $\Omega$ be the system of all forms in $w$, the $u$'s and $y$'s, which vanish for those solutions of $\Lambda$ for which $B\neq 0$.* We shall prove that $\Omega$ is irreducible.

Let $P$ and $Q$ be forms such that $PQ$ holds $\Omega$. For $s$ appropriate, $B^sP$ minus a linear combination of $Bw-A$ and its derivatives, is a form $R$ free of $w$. We obtain similarly, from a $B^tQ$, a form $S$ free of $w$. Then $RS$ vanishes for every solution of $\Sigma$ with $B\neq 0$, since every such solution furnishes a solution of $\Omega$. Hence $BRS$ holds $\Sigma$, so that either $R$ or $S$ is in $\Sigma$. If $R$ is in $\Sigma$, $B^sP$ vanishes for all solutions of $\Lambda$. Hence $P$ vanishes for all solutions of $\Lambda$ with $B\neq 0$, so that $P$ is in $\Omega$. Thus $\Omega$ is irreducible.

We notice that those forms of $\Omega$ which are free of $w$ are precisely the forms of $\Sigma$.

We shall prove that $\Omega$ has a form in $w$ and the $u$'s alone.

Let $\beta_i$, $i=1, \cdots, p$, be a form of $\Sigma$ involving only $y_i$ and the $u$'s, of a minimum rank in $y_i$. Let $F_i$ have its customary significance.

For any solution of $\Omega$ with $B\neq 0$, we write

$$w = \frac{A}{B}.$$

Representing the $j$th derivative of $w$ by $w_j$, we have

$$w_j = R_j(u_1, \cdots, u_q;\ y_1, \cdots, y_p), \tag{12}$$

where $R_j$ is rational in the $u$'s, $y$'s and their derivatives, the order of the highest derivative of $y_i$ in $R_j$ not exceeding the maximum of the orders of $\beta_i$, $A$ and $B$ in $y_i$.† The denominator of each $R_j$ will be a product of powers of $B, F_1, \cdots, F_p$. Using a sufficient number of relations (12), we eliminate‡ the $y$'s and their derivatives, obtaining a relation in $w$ and the $u$'s,

$$K = 0,$$

which holds when $BF_1 \cdots F_p$ does not vanish. As $BF_1 \cdots F_p$ is not in $\Omega$, and as $\Omega$ is irreducible, $K$ must be in $\Omega$. This proves our statement.

23. We take, for $\Omega$, a basic system of forms, analogous to (9),

$$\alpha, \alpha_1, \cdots, \alpha_p \tag{13}$$

* Of course, forms in $\Omega$ may also vanish when $B=0$.

† We could depress the order in $y_i$ to the order of $\beta_i$, but this would complicate what follows.

‡ It is clear that this elimination is of a perfectly rigorous nature.

in which $\alpha$ involves only $w$ and the $u$'s, and in which $\alpha_1, \cdots, \alpha_p$ introduce in succession $y_1, \cdots, y_p$.

If $\alpha$ is not irreducible as a polypomial in the $u$'s, $w$ and their derivatives, with coefficients in $\mathfrak{F}$, we can replace it by one of its irreducible factors. We assume, therefore, that $\alpha$ *is an irreducible form.*

We are going to prove that $\alpha_1, \cdots, \alpha_p$ are of order zero in $y_1, \cdots, y_p$, and, indeed, that $\alpha_i$ is of the first degree in $y_i$. Thus, since $\alpha_i$ with $i>1$ must be of lower degree in $y_j$ than $\alpha_j$ with $j<i$, *each equation* $\alpha_i=0$ *will express* $y_i$ *rationally in terms of* $w$, *the* $u$'s *and their derivatives.*

The determination of the content of $\Sigma$ will, in this way, be made to depend on the determination of the general solution of $\alpha=0$, which equation will be called a *resolvent* of $\Sigma$.

It is hardly necessary to call attention to the analogy which the introduction of $w$ creates, between the content of $\Sigma$, and a system of $p$ algebraic functions of $q$ variables.

Suppose that $\alpha_1$ is of order higher than zero in $y_1$. Consider any regular solution of (13) for which $BG\neq 0$. By the final remark of §20, such regular solutions exist. Without changing $w$ or the $u$'s, in the solution, we can alter the initial conditions for $y_1$ slightly, obtaining a second regular solution of (13) with $BG\neq 0$. That is, we can solve $\alpha_1=0$ for $y_1$ with the modified initial conditions, substitute the resulting $y_1$ into $\alpha_2$, solve $\alpha_2=0$ with the same initial conditions for $y_2$ which obtained in the first regular solution, and continue, determining each $y_i$. Thus, we would have two distinct solutions of $\Omega$, with the same $u$'s, with $BG\neq 0$, and with the same $w$.

Hence, $\alpha_1$ is of zero order in $y_1$. Similarly, every $\alpha_i$ is of zero order in $y_i$. Furthermore, as $\alpha_i$ is of lower rank in $y_j$ than $\alpha_j$, for $j<i$, $\alpha_i$ is of zero order in $y_j$ for $j\leqq i$.

We shall now prove that every $\alpha_i$ is linear in $y_i$.

We start with $\alpha_p$. Suppose that $\alpha_p$ is not linear in $y_p$.

Let $F_i=\partial\alpha_i/\partial y_i$ and let $H_i$ be the coefficient of the highest power of $y_i$ in $\alpha_i$.

By the familiar process of reduction, we can obtain from $B$ a form $B_1$, involving $w$, not in $\Omega$, of lower rank than each $\alpha_i$ in $y_i$ and of lower rank than $\alpha$ in $w$, such that any regular solution of (13), which causes either of the forms $B$ or $B_1$ to vanish, causes the other to vanish.

If we can show that the system

$$\alpha, \alpha_1, \cdots, \alpha_{p-1} \tag{14}$$

has a regular solution* $w, y_1, \cdots, y_{p-1}$ for which $\alpha_p$ has two distinct solutions in $y_p$ with $F_pH_pB_1G \neq 0$, we shall have forced a contradiction.

If we cannot get two distinct solutions of $\alpha_p$ of this type, it must be that for every regular solution of (14) with $H_p \neq 0$, $\alpha_p$ has a solution for which $F_pB_1G$ vanishes.†

Dividing‡ $F_pB_1G$ by $\alpha_p$, we obtain a form $\beta$, not in $\Omega$, of zero order in the $y$'s, and of lower degree than $\alpha_p$ in $y_p$, such that every common solution of $F_pB_1G$ and $\alpha_p$ is a solution of $\beta$.

Of all forms not in $\Omega$, of zero order in the $y$'s, which are of lower degree than $\alpha_p$ in $y_p$, and which, for every regular solution of (14) with $H_p \neq 0$, have a solution for $y_p$ in common with $\alpha_p$, let $\gamma$ have a minimum degree in $y_p$. Then $\gamma$ must be at least of the first degree in $y_p$, else $H_p\gamma$ would vanish for all regular solutions of (13) and would be in $\Omega$.

Let $K$ be the coefficient of the highest power of $y_p$ in $\gamma$. Then $K$ is not in $\Omega$. For $m$ appropriate,

$$K^m\alpha_p = \delta\gamma + \eta,$$

with $\delta$ of lower degree than $\alpha_p$ in $y_p$, and $\eta$ of lower degree than $\gamma$ in $y_p$. Every common solution of $\alpha_p$ and $\gamma$ makes $\eta$ vanish. Then $\eta$ must be in $\Omega$.

Thus $\delta\gamma$ is in $\Omega$, so that $\delta$, which is not zero, is in $\Omega$. Since $K^m\alpha_p$ is of higher degree in $y_p$ than $\eta$, the coefficient of the highest power of $y_p$ in $K^m\alpha_p - \eta = \delta\gamma$ is not in $\Omega$. Then the coefficient of the highest power of $y_p$ in $\delta$ is not in $\Omega$. Thus, reducing $\delta$ with respect to $\alpha_{p-1}, \cdots, \alpha$, by the familiar method, we would obtain from $\delta$ a form in $\Omega$, not zero, of lower degree than every $\alpha_i$ in $y_i$ and of lower rank than $\alpha$ in $w$.

This contradiction proves that $\alpha_p$ is linear in $y_p$.

We now consider $\alpha_{p-1}$, assuming that it is not linear in $y_{p-1}$. Since $B_1$ is of lower degree than $\alpha_p$ in $y_p$, $B_1$ is free of $y_p$. It must be that, for every regular solution of

$$\alpha, \alpha_1, \cdots, \alpha_{p-2} \tag{15}$$

with $H_{p-1} \neq 0$, $\alpha_{p-1}$ has a solution which causes $F_{p-1}H_pB_1G$ to vanish. The proof continues as for $\alpha_p$.

In dealing with $\alpha_{p-2}$, we consider that both $B_1$ and $H_p$ are free of $y_{p-1}$. The proof continues as above.

Thus every $\alpha_i$ is linear in $y_i$, and *each $y_i$ has an expression rational in in $w, u_1, \cdots, u_q$ and their derivatives, with coefficients in $\mathfrak{F}$.*

---

* A solution with $FF_1 \cdots F_{p-1}HH_1 \cdots H_{p-1} \neq 0$.

† Because (13) has regular solutions, (14) has regular solutions with $H_p \neq 0$.

‡ After a multiplication by a power of $H_p$.

24. We propose to determine which solutions of (13) other than the regular solutions are solutions of $\Omega$.

We notice first that *if*

$$u_1, \cdots, u_q;\ w;\ y_1, \cdots, y_p$$

*is a solution of* $\Omega$, *then*

$$u_1, \cdots, u_q;\ w$$

*belongs to the general solution of* $\alpha$.

For, if a form $K$ in the $u$'s and $w$ vanishes for every solution in the general solution of $\alpha$, then $K$ vanishes for every regular solution of (13) and so is in $\Omega$.

The question then arises as to which solutions of (13), for which $H_1 \cdots H_p$ vanishes, are solutions of $\Omega$.

This question is settled by the method of §§17, 18. One sees that *for a solution*

$$u_1, \cdots, u_q;\ w;\ y_1, \cdots, y_p \tag{16}$$

*of* (13) *to be a solution of* $\Omega$, *it is necessary and sufficient that in every area in which the functions of* (16) *are analytic, a point a exist such that, for every positive integer m, and for every* $\epsilon > 0$, *there is a regular solution of* (13) *in which the values of the functions and their first m derivatives at a differ from the corresponding values for* (16) *by quantities less than* $\epsilon$ *in modulus.*

It follows, as in §18, that *if a form with coefficients meromorphic in* $\mathfrak{A}$, *the coefficients not belonging necessarily to* $\mathfrak{F}$, *vanishes for all regular solutions of* (13), *the form vanishes for all solutions of* $\Omega$.

25. We shall derive a result which is, to some extent, a converse of the result of §23.

Suppose that we have a differential equation

$$\alpha(u_1, \cdots, u_q;\ w) = 0, \tag{17}$$

$\alpha$ being an irreducible form, with coefficients in $\mathfrak{F}$.

Let there be given $p$ rational combinations of $u_1, \cdots, u_q;\ w$ and their derivatives, with coefficients in $\mathfrak{F}$,

$$y_i = \frac{P_i(w;\ u_1, \cdots, u_q)}{Q_i(w;\ u_1, \cdots, u_q)} \qquad (i = 1, \cdots, p), \tag{18}$$

no $Q_i$ vanishing for every solution in the general solution of $\alpha$.

Consider any regular solution of (17) and (18), that is, a set

$$u_1, \cdots, u_q;\ w;\ y_1, \cdots, y_p$$

consistent with (17), (18), in which $u_1, \cdots, u_q; w$ is a regular solution of (17), and for which, naturally, no $Q_i$ vanishes. It can be shown, as in the preceding sections, that there is, for every $i$, a form in $y_i$ and the $u$'s which vanishes for all regular solutions.

Consider the system $\Sigma$ of all forms in the $y$'s and $u$'s which vanish for all regular solutions of (17), (18).

We shall prove that $\Sigma$ is irreducible.

Let $RS$ hold $\Sigma$. If we substitute (18) into $R$, $R$ becomes a rational combination of $u_1, \cdots, u_q; w$ and their derivatives

$$\frac{T(w; u_1, \cdots, u_q)}{U(w; u_1, \cdots, u_q)},$$

where $U$ is a product of powers of the $Q_i$'s. Similarly, $S$ becomes a rational combination $V/W$ of the $u$'s, $w$ etc.

For the $w$ and $u$'s of any regular solution of (17), (18), $TV$ vanishes. This means that, for every regular solution of $\alpha$,

$$TVQ_1 \cdots Q_p$$

vanishes. Then either $T$ vanishes for all regular solutions of $\alpha$, or $V$ does. Hence either $R$ vanishes for all regular solutions of (17), (18), or $S$ does.

Thus $\Sigma$ is irreducible. Its content is an irreducible manifold which is contained in every manifold which contains the regular solutions of (17), (18) with $w$ suppressed.

Consider the system $\Omega$ of all forms in the $u$'s, $y$'s and $w$, which vanish for the regular solutions of (17), (18). The above discussion shows also that $\Omega$ is irreducible.

The results of this section hold even if $\mathfrak{F}$ consists purely of constants.

### Invariance of the integer $q$

26. We propose to show that *the number $q$ of arbitrary indeterminates depends only upon the system $\Sigma$ and not on the manner in which the $u$'s are selected.*

The assumption that $\Sigma$ is irreducible is essential. But it must be realized, in this connection, that §19 develops the idea of arbitrary indeterminate in a rather special way.*

* Consider the system of equations $u_1y_1=u_2y_2=u_3y_3=0$. These equations imply no relations either among the $u$'s or among the $y$'s. Still each $u$ appears in a form with $y$'s alone, and each $y$ appears with $u$'s alone.

It will suffice to prove that, given any $q+1$ indeterminates among the $u$'s and $y$'s,

$$z_1, \cdots, z_{q+1},$$

there exists a form of $\Sigma$ which involves only the $z$'s.

Let us suppose that $\mathfrak{F}$ does not consist purely of constants, and let us consider the regular solutions of (13). For $u_1, \cdots, u_q; w$ in such a solution, (13) gives a rational expression for each $z_i$. If a $z_i$ happens to be a $u$, say $u_j$, the expression for $z_i$ is simply $u_j$. We write

$$z_i = R_i(w; u_1, \cdots, u_q) \qquad (i = 1, \cdots, q+1). \tag{19}$$

On differentiating (19) repeatedly, we get expressions for the derivatives of the $z$'s which are rational in terms of the $u$'s, $w$ and their derivatives. Making use of the relation $\alpha=0$, we transform these expressions so as not to contain derivatives of $w$ of order higher than $r$, the order of the resolvent in $w$.

None of the expressions thus obtained will have a denominator which vanishes for $u$'s and $w$ in a regular solution of (13).

Now, if we differentiate the $q+1$ relations (19) often enough, the $z$'s and their derivatives will become more numerous than the $u$'s, their derivatives and $w, \cdots, w_r$.

It follows that there exists a polynomial in the $z$'s and their derivatives, with coefficients in $\mathfrak{F}$, which vanishes for all regular solutions of (13). The form thus obtained belongs to $\Sigma$.

Suppose now that $\mathfrak{F}$ contains only constants. Let $\mathfrak{F}_1$ be the field obtained from $\mathfrak{F}$ by the adjunction of $x$. Let $\Sigma_1, \cdots, \Sigma_s$ be irreducible systems in $\mathfrak{F}_1$ such that $\Sigma$ holds each of them and that every solution of $\Sigma$ is a solution of one of them.*

Suppose that $\Sigma$ has, in $\mathfrak{F}$, two sets of arbitrary indeterminates, $u_1, \cdots, u_q$ and $z_1, \cdots, z_t$ with $t \neq q$. We are going to arrive at the contradiction that both the $u$'s and the $z$'s are arbitrary for some $\Sigma_i$ in $\mathfrak{F}_1$.

Suppose that this is not so, and that each $\Sigma_i$ has either a form in the $u$'s alone or a form in the $z$'s alone. Then the product of $s$ such forms, one from each $\Sigma_i$, will vanish for every solution of $\Sigma$.

Consider then any form $K$, taken from some $\Sigma_i$, which is a polynomial in $x$, the $u$'s and their derivatives, with coefficients in $\mathfrak{F}$. Let $K$ be irreducible as a polynomial in $x$ etc.† Let $K'$ be the derivative of $K$. Then the resultant of $K$ and $K'$ with respect to $x$, which is not zero, vanishes for any $u$'s which

* Whether $\Sigma$ can be reducible in $\mathfrak{F}_1$ is a question.

† Irreducibility may certainly be assumed for the $s$ forms considered above.

make $K$ vanish. The resultant is a form in the $u$'s, with coefficients in $\mathfrak{F}$.

Thus, there is a product of $s$ forms, some in the $u$'s alone, some in the $z$'s alone, with coefficients in $\mathfrak{F}$, which holds $\Sigma$. This cannot be, as $\Sigma$ is irreducible in $\mathfrak{F}$. Thus, there is a $\Sigma_i$ for which both the $u$'s and the $z$'s are arbitrary. This completes the proof.

### Invariance of order of resolvent

27. We propose to show that, $u_1, \cdots, u_q$ *being selected, the order with respect to* $w$ *of the resolvent is independent of the choice of* $w$.

Having taken a definite $w$, and having formed the resolvent in $w$, $\alpha = 0$, let us form a second rational combination of the $u$'s, $y$'s and their derivatives,

$$v = \frac{C}{D}, \tag{20}$$

with $v$ meeting all specifications placed above on $w$. Let the resolvent in $v$ be $\beta = 0$.

Since $D$ is not in $\Sigma$, $D$ is not in the system $\Omega$ based on $w$. Hence, there are regular solutions of (13) for which $D \neq 0$. Taking any such regular solution, let the $y$'s in it, expressed in terms of $w$ and the $u$'s, be substituted into the expression (20) for $v$. We find, for $v$, an expression in the $u$'s, $w$ and their derivatives,

$$v = R(w; u_1, \cdots, u_q). \tag{21}$$

Using the equation $\alpha = 0$, if necessary, we may suppose that $R$ involves no derivatives of $w$ of order higher than $r$, the order of $\alpha$ in $w$.

We differentiate (21) $r$ times, and find that, $v_i$ being the $i$th derivative of $v$,

$$v_i = R_i(w; u_1, \cdots, u_q) \qquad (i = 1, \cdots, r), \tag{22}$$

each $R_i$ being rational, and involving no derivative of $w$ beyond the $r$th. From (21), (22) and $\alpha = 0$, we can eliminate $w, \cdots, w_r$, and obtain an algebraic relation

$$K(u_1, \cdots, u_q; v) = 0, \tag{23}$$

at most of order $r$ in $v$.

The relation (23) holds for any $v$ given by (20), if the $u$'s and $y$'s, for which $D \neq 0$, belong to a regular solution of (13). Now, if we replace $v$ in (23) by its expression (20), (23) becomes a relation in the $u$'s and $y$'s,

$$\frac{L(u_1, \cdots, u_q; y_1, \cdots, y_p)}{D^m} = 0,$$

with $m$ a positive integer. Then $L$ vanishes for all $u$'s and $y$'s in a regular solution of (13), for which $D \neq 0$. Thus $DL$ is in $\Omega$, and as $D$ is not in $\Omega$, $L$ is in $\Omega$. Hence $L$ is in $\Sigma$, and (23) holds for any $v$ given by (20), where the $u$'s and $y$'s are any solution of $\Sigma$ with $D \neq 0$.

If, then, $\Omega'$ is the system associated with $v$ as $\Omega$ is with $w$, $K$, in (23), is in $\Omega'$. This proves that the order of $\beta$ in $v$ does not exceed the order of $\alpha$ in $w$. From considerations of symmetry, it follows that the two orders are equal. This proves our statement.

The degree of the resolvent in the highest derivative of $w$ does depend on the manner of choosing $w$. Consider, for instance, the system, irreducible in the field of all rational functions,

$$\frac{d}{dx}y_1 = 1, \quad y_2 = y_1^2 .$$

As the solution of the system is $y_1 = x + a$, $y_2 = (x+a)^2$, we may evidently take $w = y_1$. The resolvent becomes $dw/dx = 1$. On the other hand, if we take $w = y_1 + y_2$, the resolvent becomes of the second degree in $dw/dx$.

The order of the resolvent depends on the choice of the $u$'s. For instance

$$\frac{dy_1}{dx} - y_2 = 0$$

is an irreducible system in the field of rational functions. If we let $u_1 = y_2$, we get a resolvent of the first order. If we let $u_1 = y_1$, we get a resolvent of zero order.

### Adjunction of new functions to $\mathfrak{F}$

28. Assuming $\mathfrak{F}$ not to consist purely of constants, we shall study the circumstances under which $\Sigma$ can become reducible through the adjunction of new functions to $\mathfrak{F}$. The adjoined functions are assumed to be meromorphic in $\mathfrak{A}$.

We form a resolvent $\alpha = 0$ for $\mathfrak{F}$, using a $w$ whose denominator, $B$, is unity.

Suppose that the irreducible factors of $\alpha$, in the enlarged field, $\mathfrak{F}_1$, are $\beta_1, \cdots, \beta_s$.

Then, by §25, for each $j$ from 1 to $s$, the system of equations

$$\beta_j = 0, \ \alpha_1 = 0, \cdots, \alpha_p = 0, \tag{24}$$

where the $\alpha$'s are those of (13), defines a system $\Sigma_j$ of forms in the $u$'s and $y$'s, with coefficients in $\mathfrak{F}_1$, $\Sigma_j$ being irreducible in $\mathfrak{F}_1$.

We shall prove that $\Sigma$ holds every $\Sigma_j$, that no $\Sigma_h$ holds any $\Sigma_k$ with $k \neq h$,

and that every solution of $\Sigma$ is a solution of some $\Sigma_j$. *Thus, the systems $\Sigma_j$ will furnish the resolution of $\Sigma$ into irreducible systems, in $\mathfrak{F}_1$.*

Every regular solution of $\alpha$ is a regular solution of some $\beta_j$. First, in no $\beta_j$ can the coefficient of the highest power of $w_r$* vanish for a regular solution of $\alpha$. Again, since

$$\frac{\partial \alpha}{\partial w_r} = \beta_2 \cdots \beta_s \frac{\partial \beta_1}{\partial w_r} + \cdots,$$

$\partial \beta_j / \partial w_r$ cannot vanish for a regular solution of $\alpha$ if $\beta_j$ does. Thus every regular solution of (13) is a regular solution of some system (24). Hence a solution of $\Sigma$ obtained by suppressing $w$ in a regular solution of (13) is a solution of some $\Sigma_j$. Since $B=1$, every solution of $\Sigma$ is obtained from some solution of $\Omega$.

Suppose now that some solution of $\Sigma$ is not a solution of any $\Sigma_j$. Let $C_j$ be a form of $\Sigma_j, j=1, \cdots, s$, which does not vanish for the solution. Then $C_1 \cdots C_s$ does not vanish for the solution. This contradicts the final remark of §24. Hence every solution of $\Sigma$ is a solution of some $\Sigma_j$.

Let $\Omega_j (j=1, \cdots, s)$ be the system of all forms in $w$, the $u$'s and $y$'s, with coefficients in $\mathfrak{F}_1$, which vanish for all regular solutions of (24). As was seen in §25, $\Omega_j$ is irreducible.

Let $H$ be the coefficient of the highest power of $w_r$ in $\alpha$. If $H$ were in some $\Omega_j$, it would vanish for all regular solutions of $\beta_j$. This cannot be, for $H$ is of order less than $r$ in $w$.

If $F = \partial \alpha / \partial w_r$ were of order $r$ and were in some $\Omega_j$, it would be divisible by $\beta_j$. Then $\alpha$ would be reducible in $\mathfrak{F}$.

Consider any form $P$ of $\Omega$. Any regular solution of (24), for any $j$, for which $FH$ does not vanish, causes $P$ to vanish. Hence $FHP$ is in $\Omega_j$, so that $P$ is in $\Omega_j$. Thus every form of $\Sigma$ is in $\Sigma_j$, so that $\Sigma$ holds every $\Sigma_j$.

The foregoing shows also that $\Omega$ holds every $\Omega_j$. Thus every $\Omega_j$ contains the form $w-A$ used in building $\Omega$. It follows easily that $\Omega_j$ holds and is held by the system $\Lambda_j$ obtained by adjoining $w-A$ to $\Sigma_j$.

This means that if $\Sigma_h$ held $\Sigma_k$, where $k \neq h$, then $\Omega_h$ would hold $\Omega_k$. Then $\beta_h$ would be in $\Omega_k$, and would be divisible by $\beta_k$. This would make $\alpha$ reducible in $\mathfrak{F}$. Thus no $\Sigma_h$ can hold a $\Sigma_k$ with $k \neq h$.

Thus, *for $\Sigma$ to be reducible in $\mathfrak{F}_1$, it is necessary and sufficient that the resolvent of $\Sigma$ relative to $\mathfrak{F}$ be algebraically reducible in $\mathfrak{F}_1$.*†

29. The question might be asked as to whether $\Sigma$, irreducible in $\mathfrak{F}$ for

* $\alpha$ of order $r$ in $w$.

† We recall the assumption that $B=1$.

the area $\mathfrak{A}$, can be reducible in $\mathfrak{F}$ for some area $\mathfrak{B}$ contained in $\mathfrak{A}$. We shall show that the answer is negative.

We begin by showing that if a form $K$, with coefficients in $\mathfrak{F}$, vanishes for all solutions of $\Sigma$ analytic in a part of $\mathfrak{B}$, then $K$ vanishes for all solutions of $\Sigma$. Suppose then that $K$ is not in $\Sigma$. By the familiar process of reduction, we obtain from $K$ a form $L$, in $w$ and the $u$'s, of lower rank in $w$ than $\alpha$, which vanishes for every regular solution of $\alpha$, analytic in a part of $\mathfrak{B}$, for which no $H_i$ vanishes.* As in §15, we reach the absurdity that

$$LFHH_1 \cdots H_p$$

is divisible by $\alpha$.†

Now if $P$ and $Q$ are two forms, with coefficients in $\mathfrak{F}$, such that $PQ$ vanishes for all solutions of $\Sigma$ analytic in a part of $\mathfrak{B}$, than $PQ$ vanishes for all solutions analytic in any part of $\mathfrak{A}$. This means that either $P$ or $Q$ is in $\Sigma$, so that $\Sigma$ is irreducible in $\mathfrak{B}$.

---

* $H_i$ is the coefficient of $y_i$ in $\alpha_i$.

† When $\mathfrak{F}$ contains only constants, we adjoin $x$ to $\mathfrak{F}$ and consider the irreducible systems into which $\Sigma$ decomposes.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# THE FOUNDATIONS OF A THEORY OF THE CALCULUS OF VARIATIONS IN THE LARGE IN $m$-SPACE (SECOND PAPER)*

BY

MARSTON MORSE

**1. Introduction.** The general problem† is the classification and existence of extremals under all kinds of conditions, a non-linear boundary value problem in the large. This paper deals primarily with the fixed end point problem.

The first paper by the author dealt with the fixed end point problem in the small, and with the problem with one end point variable in general. The latter problem could be readily studied in the large because it corresponds to the problem of finding the critical points of a function whose critical points are in general isolated.

Such is not the case in the fixed end point problem. The critical sets appearing there take the form of $n$-dimensional loci. This extreme difficulty is met by the aid of deformations. The author's characterization of the number of conjugate points on an extremal by means of deformations is essential. See Morse II.

A second difficulty which long appeared insurmountable was the fact

* Presented to the Society as part of an address given at the request of the program committee April 6, 1928, under the title *The critical points of functions and the calculus of variations in the large*; see also Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 38–54; received by the editors May 8, 1930.

† The following references should be made.

For work on the absolute minimum including work of Hilbert see Bolza, *Vorlesungen über Variationsrechnung*, 1909, pp. 419–437, and Tonelli, *Fondamenti di Calcolo delle Variazioni*, vol. 2. Further references will be found in these works.

Signorini, Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 187–193. Here broken geodesics are used in the quest for minimizing periodic orbits.

Birkhoff, *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9. In chapter V, Birkhoff makes new and effective use of deformations of broken geodesics to obtain basic periodic motions. His coupling of the minimum and minimax methods was undoubtedly the first step towards a general theory.

The following papers by the author will be referred to: I These Transactions, vol. 27 (1925), pp. 345–396. II These Transactions, vol. 30 (1928), pp. 213–274. III These Transactions, vol. 31 (1929), pp. 379–404. IV Mathematische Annalen, vol. 103 (1930), pp. 52–69.

See also the following abstract: MM. Lusternik et Schnirelmann, *Existence de trois géodésiques fermées sur toute surface de genre* 0, Comptes Rendus, vol. 188 (1929), No. 8, p. 534.

For references to Bliss, Mason, Carathéodory, and Hadamard, see the papers Morse II and III.

that the domains of definition of the functions whose critical points were sought were not complexes, or at least could not readily be reduced to finite complexes.

This difficulty was again surmounted with the aid of deformations. The regions involved were shown to be deformable on themselves into subregions which were complexes. The existence of these deformations was a matter of the calculus of variations, their use a matter of analysis situs. Both aspects were indispensable.

A final difficulty as well as source of great interest was the fact that the domains usually have infinite connectivities. From the point of view of the calculus of variations this involved a study of the density of infinite sets of conjugate points of a fixed point. Over against the infinite set of connectivities is set a conjugate number sequence belonging to the calculus of variations, and the interrelations of these two infinite sequences appear central.

As a particular example, one can prove the existence of infinitely many geodesics joining any two points on any regular topological image of an $m$-sphere.

The author wishes to state in passing that a considerable part of the analysis of this problem has been carried over to the problem of periodic extremals.* The latter problem, however, presents additional difficulties both from the point of view of the calculus of variations and that of analysis situs. It involves the study of homologies on a given complex among complexes which are restricted to product complexes. The simplest complexes present decidedly new aspects from this point of view.

## I. The problem and its deformations

**2. The hypotheses.** Let $S$ be a closed region representable as an $m$-dimensional complex† $A_m$ in the space of the variables $(x_1, \cdots, x_m) = (x)$. Let

$$F(x_1, \cdots, x_m, r_1, \cdots, r_m) = F(x, r) \tag{2.1}$$

be a *positive, analytic* function of its arguments for $(x)$ on a region slightly larger than $S$, and for $(r)$ any set not (0). Suppose further that $F$ is positively homogeneous of order one in the variables $(r)$. We employ the parametric form, taking $F(x, \dot{x})$ as our integrand and

$$J = \int_{t_1}^{t_2} F(x, \dot{x})dt$$

* See Marston Morse, *Closed extremals*, Proceedings of the National Academy of Sciences, vol. 15 (1929), p. 856.

† See Veblen, The Cambridge Colloquium, 1916, Part II, *Analysis Situs*. Terms in analysis situs will be used in the sense of Veblen unless otherwise defined.

as our integral, where $(\dot{x})$ stands for the set of derivatives of $(x)$ with respect to the parameter $t$. We suppose that the problem is *positively regular*, that is, that

$$\sum_{ij} F_{r_i r_j}(x, r)\eta_i\eta_j > 0 \qquad (i, j = 1, 2, \cdots, m) \tag{2.2}$$

for $(x)$ and $(r)$ as before, and $(\eta)$ any set not (0) nor proportional to $(r)$.

*We suppose that S's boundary is extremal-convex.**

That is, we suppose there exists a positive constant $e$ so small that any extremal segment on which $J<e$ and which joins two boundary points of $S$ will lie interior to $S$ except at most for its end points.

Let $P$ and $Q$ be any two distinct points on $S$. Let there be given an extremal segment $g$ which joins $P$ to $Q$ on $S$. Let $\lambda$ be the direction of $g$ at $P$. Let the direction cosines of the directions neighboring $\lambda$ be regularly† represented as functions of $m-1$ parameters $(\alpha)$. Let $s$ be the value of $J$ taken along the extremal through $P$ in the direction determined by $(\alpha)$. On the extremals issuing from $P$ with the directions determined by $(\alpha)$ the coordinates $(x)$ will be analytic functions of $s$ and $(\alpha)$.

*All of the extremal segments through P neighboring g cannot go through Q.*

This could happen if $S$ were a closed manifold, but it cannot happen if $S$ is an extremal-convex region. For if all such extremals did go through $Q$ one could prove first that $s$ would be a constant at $Q$, and secondly, one could show by a process of analytic continuation that all extremals whatsoever through $P$ would reach $Q$ at least as soon as they reach the boundary.

This cannot occur. For in particular the class of extremals which join $P$ to an arbitrary point $R \neq Q$ of the boundary, and give an absolute minimum to $J$ between $P$ and $R$, cannot all pass through $Q$ prior to $R$ without violating their minimizing property. Thus the statement in italics is proved.

If in terms of the parameters $s$ and $(\alpha)$ one writes down the $m$ conditions that the extremal through $P$ pass through $Q$, one sees from the theory of analytic functions that the extremals which join $P$ to $Q$ on $S$ with $J$ less than a finite constant $J_0$, are conditioned as follows:

*These extremals are either finite in number or else make up a finite set of families on which $(x)$ is representable by means of the parameters $s$ and $(\alpha)$,*

---

* Instead of assuming that $S$ was an extremal-convex region in $m$-space, we could have assumed that $S$ was a regular $(m-1)$-dimensional manifold in $m$-space. Such a change would entail at most obvious changes in the following. The results for this case will be reviewed at the end of the paper.

† That is, the direction cosines are to be analytic functions of the parameters $(\alpha)$ of such sort that not all of the jacobians of $m-1$ of the direction cosines with respect to the parameter $(\alpha)$ are zero.

*with $s$ a constant for each family, and $(\alpha)$ analytic in general for each family on a suitably chosen "Gebilde"* of $r$ independent variables with $0<r<m-1$.*

We note the following.

*In the plane there are at most a finite number of extremals on $S$ through $P$ and $Q$ on which $J<J_0$.*

We wish to call attention to the case where $Q$ is not conjugate to $P$ on any of the extremals joining $P$ to $Q$ on which $J<J_0$. This is the general case. In this case there are at most a finite number of extremals joining $P$ to $Q$ on which $J<J_0$, as is readily seen from the fact that each such extremal is isolated.

**3. The region $\Sigma$ and function $J(\pi)$.** Instead of considering the set of all curves joining $P$ to $Q$ our purposes will be equally well served by considering a particular class of broken extremals joining $P$ to $Q$. To proceed we shall give successively a number of definitions.

We shall call the value of $J$ taken along a curve $\gamma$ the *$J$-length* of $\gamma$.

*The constant $J_0$.* We shall restrict ourselves to curves of $J$-length less than $J_0$, where $J_0$ is a positive constant larger than the absolute minimum of $J$ along curves joining $P$ to $Q$, and where further $J_0$ is not equal to the value of $J$ along any extremals joining $P$ to $Q$.

*The constant $\rho$.* It is well known that there exists a positive constant $e_1$ small enough to have the following properties. Any extremal $E$ lying on $S$ and with a $J$-length less than $e_1$, will give an absolute minimum to $J$ relative to all curves of class $D'$ joining its end points. The coördinates $(x)$ of any point on $E$ will be analytic functions of the coördinates of the end points of $E$ and of the distance of $(x)$ from the initial end point $O$ of $E$, at least as long as $E$ does not reduce to a point. The set of all extremal segments issuing from $O$ with $J$-lengths equal to $e_1$ will form a field covering a neighborhood of $O$ in a one-to-one manner, $O$ alone excepted. With $e_1$ thus chosen we now choose the positive constant $\rho$ so that

$$\rho < e_1, \quad \rho < e, \tag{3.1}$$

where $e$ is the constant used in the definition of the term extremal-convex.

Any extremal segment on $S$ whose $J$-length is at most $\rho$ will be called an *elementary extremal.*

*Admissible broken extremals.* Let $n$ be a positive integer so large that

$$J_0 < (n+1)\rho. \tag{3.2}$$

Consider now the class of broken extremals $g$ which satisfy the following conditions:

* See Osgood, *Funktionentheorie*, II.

I. The broken extremal joins $P$ to $Q$ on $S$.

II. It consists of $n+1$ elementary extremals.

III. Its $J$-length is less than $J_0$.

Such a broken extremal $g$ will be termed *admissible.*

*Admissible points* $(\pi)$. Let

$$P, P_1, \cdots, P_n, Q \tag{3.3}$$

be the successive ends of the elementary extremals of $g$. We admit the possibility that two or more of these points be coincident. Let $(\pi)$ represent the set of $mn$ variables which give the coördinates of the points

$$P_1, P_2, \cdots, P_n, \tag{3.4}$$

taking first the coördinates of $P_1$, then those of $P_2$, and so on. The points (3.4) will be called the vertices of $g$. A point $(\pi)$ derived from the vertices of an admissible broken extremal will itself be called *admissible.*

There are infinitely many admissible points $(\pi)$. In particular let $g_1$ be an extremal joining $P$ to $Q$ with a $J$-length less than $J_0$. If we divide $g_1$ into $n+1$ successive extremals of equal $J$-length, these extremals turn out to be elementary extremals by virtue of (3.2), so that the resulting $n$ points of division of $g_1$ combine into an admissible point $(\pi)$. This point $(\pi)$ has the further property that any point $(\pi)$ in its neighborhood will also be admissible.

*The set of all admissible points* $(\pi)$ *will form an* $mn$*-dimensional domain* $\Sigma$.

*The value of the integral* $J$ *taken along the broken extremal determined by an admissible point* $(\pi)$ *from the point* $P$ *to the point* $Q$ *will be denoted by* $J(\pi)$.

**4. The boundary of $\Sigma$.** Corresponding to an admissible point $(\pi)$, let $M(\pi)$ be the maximum of the $J$-lengths of the elementary extremals that make up the broken extremal determined by $(\pi)$.

The boundary of $\Sigma$ will consist of points $(\pi)$ of one or more of the following types:

Type I: Points at which $J(\pi)=J_0$;

Type II: Points at which $M(\pi)=\rho$;

Type III: Points corresponding to which at least one vertex $P_i$ lies on the boundary of $S$.

The function $J(\pi)$ is analytic on $\Sigma$ at least as long as successive vertices remain distinct. A point $(\pi)$ corresponding to which the successive vertices are distinct will be called a *critical point* of $J(\pi)$ if all of the first partial derivatives of $J(\pi)$ are zero at that point. We shall prove for the case of distinct vertices that $J(\pi)$ will have a critical point $(\pi)$ when and only when

the corresponding broken extremal reduces to an unbroken extremal $\gamma$ joining $P$ to $Q$.

For the partial derivative of $J(\pi)$ with respect to the $i$th coördinate of a vertex $(x)$, in the case of distinct vertices, is seen to be

$$F_{r_i}(x, p) - F_{r_i}(x, q) \qquad (i = 1, 2, \cdots, m) \tag{4.1}$$

where $(p)$ and $(q)$ give the directions at $(x)$ of the elementary extremals preceding and following $(x)$ respectively. If the differences (4.1) all vanished we would have

$$\sum_i [p_i F_{r_i}(x, p) - p_i F_{p_i}(x, q)] = 0. \tag{4.2}$$

But the sum (4.2) equals the Weierstrass $E$-function $E(x, p, q)$ which is known not to vanish* for $(p) \neq (q)$ whenever the hypothesis of regularity of §2 is granted. Hence $(p) = (q)$ if $J(\pi)$ has a critical point. The corresponding extremal can therefore have no corners.

If account be taken of the freedom in moving vertices on an extremal $\gamma$ joining $P$ to $Q$, it is seen that each such extremal corresponds to an $n$-dimensional set of critical points $(\pi)$.

**5. The deformation $T$.** The boundary of $\Sigma$ is exceedingly complex. We can avoid examining its structure more closely by showing how to deform $\Sigma$ continuously as a whole into a part of $\Sigma$ which contains none of the boundary points of $\Sigma$ of types I and II. These deformations amount to deformations of broken extremals. For future use they need to have the following properties.

($\alpha$) *They do not increase $J(\pi)$ or $M(\pi)$ beyond their initial values.*

($\beta$) *They deform admissible broken extremals into admissible broken extremals.*

($\gamma$) *They deform continuous families of broken extremals continuously.*

*Such deformations will be called J-deformations.*

*The deformation $D_1$.* On an admissible broken extremal $g$ let a point $U$ be given. Let the value of $J$ taken along $g$ from $P$ to $U$ be termed the *J-coördinate*, $u$, of the point $U$ on $g$. In case $u$ is a function of the time $t$ it will be convenient to term $du/dt$ the *J-rate* of $U$ on $g$.

As the time $t$ varies from 0 to 1 let the $n$ vertices $P_i$ on $g$ move along $g$ from their initial positions to a set of positions which divide $g$ into $n+1$ successive segments of equal $J$-lengths, each vertex moving at a constant $J$-rate.

We term this the deformation $D_1$. It is readily seen that $D_1$ is a $J$-deformation. We also note the following.

* See Bliss, *The Weierstrass E-function for problems of the calculus of variations in space*, these Transactions, vol. 15 (1914), pp. 369–378.

LEMMA 5.1. *The deformation $D_1$ carries each admissible broken extremal into one for which*

$$M(\pi)(n+1) < J_0.$$

*The deformation $D_2$.* Let $h_i$ be the $i$th elementary extremal of $g$. As the time $t$ increases from 0 to 1 let points $H_i(t)$ and $K_{i+1}(t)$ start from $P_i$ and move away from $P_i$, respectively on $h_i$ and $h_{i+1}$, at $J$-rates equal to one half the $J$-lengths of $h_i$ and $h_{i+1}$. Let $H_i(t)$ and $K_{i+1}(t)$ be joined by an elementary extremal $E_i(t)$. The deformation $D_2$ is hereby defined as one in which $P_i$ is replaced for each $t$ by the point $P_i(t)$ which divides the elementary extremal $E_i(t)$ into two segments of equal $J$-lengths.

To show that $D_2$ is a $J$-deformation we shall first show that $M(\pi)$ is not increased beyond its initial value.

To that end let $h_i$ also represent the $J$-length of $h_i$. Let $g(t)$ be the broken extremal replacing $g$ at the time $t$. Note that the end points of the $i$th elementary extremal of $g(t)$ are also connected by a set of three elementary extremals joining successively the four points

$$P_{i-1}(t), \quad K_i(t), \quad H_i(t), \quad P_i(t), \tag{5.1}$$

the values $i=1$ and $n+1$ excepted. The $J$-lengths of these three elementary arcs are respectively at most

$$\frac{t}{4}(h_{i-1}+h_i), \quad (h_i - th_i), \quad \frac{t}{4}(h_i+h_{i+1}). \tag{5.2}$$

If the three constant $h_{i-1}$, $h_i$, and $h_{i+1}$ be replaced by their maximum, the sum of the three quantities in (5.2) is readily seen to be at most that maximum. Thus the deformation $D_2$ does not increase $M(\pi)$ beyond its initial value.

With this established one sees readily that the remaining properties of $J$-deformations as previously listed are satisfied by $D_2$.

The deformations $D_1$ and $D_2$ combine to give us a deformation $T$ concerning which we have the following lemma.

LEMMA 5.2. *If $\sigma$ be a closed set of admissible points $(\pi)$ at a positive distance from the set of critical points of $J(\pi)$, the product deformation $T=D_1D_2$ will carry each point $(\pi)$ of $\sigma$ into an admissible point $(\pi')$ for which*

$$J(\pi') < J(\pi) - d,$$

*where $d$ is a positive constant.*

In the region $\Sigma$ let $N$ be a closed neighborhood of the critical points of $J(\pi)$ so small that $N$ and $\sigma$ have no points in common. We divide the points $(\pi)$ of $\Sigma$ into three classes according to the nature of their images $(\pi)_1$ under $D_1$ at the time $t=1$.

Class I shall contain $(\pi)$ if $(\pi)_1$ determines a broken extremal with at least one elementary extremal $k$ of zero $J$-length.

Class II shall contain $(\pi)$ if $(\pi)_1$ is in $N$ and $(\pi)$ is not in Class I.

Class III shall contain those points of $\sigma$ which are neither in Class I nor in Class II.

The proof of the lemma will be given separately for each class.

Class I. To define $D_1$ we divided the given broken extremal into segments of equal $J$-length. When $(\pi)$ is in Class I, one of these segments, say $h$, is replaced under $D_1$ by the elementary extremal $k$ of zero $J$-length. Since the $J$-length of $h$ is not zero, $J(\pi)$ must be lessened under $D_1$.

Class II. In general the only points $(\pi)$ for which $J(\pi)$ is not lessened under $D_1$ are those for which the broken extremal determined by $(\pi)$ is identical with the one determined by $(\pi)_1$. Hence $J(\pi)$ is lessened by $D_1$ for every point $(\pi)$ of Class II.

Class III. Here the broken extremal determined by $(\pi)_1$ has no elementary extremals of zero $J$-length, and has at least two successive elementary extremals which intersect at an angle which is not straight. The application of $D_2$ to $(\pi)_1$ will then lessen $J(\pi)_1$.

Thus under $T=D_1D_2$, $J(\pi)$ will be lessened for each point of $\sigma$. Since $\sigma$ is closed, and $T$ is continuous on $\sigma$, $J(\pi)$ will be lessened on $\sigma$ by at least some definite positive constant $d$, independent of $(\pi)$ on $\sigma$. Thus the lemma is proved.

6. **The domain $(a, \rho)$ and its connectivities.** Let $a$ be a non-critical value of $J(\pi)$, and $r$ any positive constant at most the constant $\rho$ of §3.

*The set of points $(\pi)$ which satisfy*

$$J(\pi) < a,\ M(\pi) < r,\ a < J_0 \tag{6.1}$$

*will be called a domain $(a, r)$.*

We shall be more particularly concerned with domains $(a, r)$ for which $r=\rho$. Because of the complicated nature of the boundary of a domain $(a, \rho)$ it is not feasible to try to break such a domain up into cells. Because of our $J$-deformations it is fortunately possible to study the topological properties of the domain $(a, \rho)$ by means of approximating complexes.

By hypothesis the region $S$ of the space of the $m$ variables $(x)$ can be broken up into $m$-cells so as to form an $m$-complex $A_m$. We provide $n$ copies of $A_m$, namely

$$A_m^1,\ A_m^2,\ \cdots,\ A_m^n, \tag{6.2}$$

upon which, in particular, we suppose the $n$ vertices, respectively, of an admissible point $(\pi)$ lie. The product complex $A_{mn}$ formed by taking an

arbitrary point from each of the complexes (6.2) will represent a portion of the space of the points $(\pi)$ of which each domain $(a, r)$ will be a subset.

By virtue of our choice of the integer $n$ in (3.2) we can choose a positive constant $e$ so small that

$$J_0 < (\rho - e)(n + 1), \tag{6.3}$$

choosing $e$ also so small that no constant on the interval between $a$ and $a-e$, including $a-e$, is a critical value of $J(\pi)$. If the deformation $T$ be applied to the domain $(a, \rho)$ it follows from (6.3) and Lemma 5.1 that the resulting points $(\pi)$ will be such that

$$M(\pi) < \rho - e.$$

It follows then from Lemma 5.2 that a sufficient number of iterations of $T$ will $J$-deform the domain $(a, \rho)$ on itself into a set of points on the domain $(a-e, \rho-e)$.

Let $C_{mn}$ be a sub-complex of cells of $A_{mn}$ that contains all of the points of $(a-e, \rho-e)$. If $A_{mn}$ be sufficiently finely divided, $C_{mn}$ may be chosen so as to consist only of points on $(a, \rho)$. It will serve as our $e$-approximation of $(a, \rho)$.

As we have seen, any complex of $A_{mn}$ on $(a, \rho)$ can be $J$-deformed on $(a, \rho)$ into a complex on $C_{mn}$. It follows that any cycle* on $(a, \rho)$ is homologous (always mod 2) on $(a, \rho)$ to a cycle of cells of $C_{mn}$. Therefore all cycles on $(a, \rho)$ are homologous to a finite number of such cycles.

By a complete *$j$-set of non-bounding* $j$-cycles on $(a, \rho)$ will be meant a set of $j$-cycles $(K)_j$ on $(a, \rho)$ with the following properties.

I. Every $j$-cycle on $(a, \rho)$ is homologous to a linear combination of members of $(K)_j$.

II. There are no proper† homologies on $(a, \rho)$ between the members of $(K)_j$.

The following is an almost immediate consequence of the definition.

*The number of $j$-cycles in a complete $j$-set for $(a, \rho)$ is the same for all such complete $j$-sets.*

*The $j$th connectivity number $R_j$ of the domain $(a, \rho)$ will now be defined as the number of $j$-cycles in a complete $j$-set.*

We come now to a first theorem.

THEOREM 1. *If $a$ and $b$, $a<b$, are two non-critical values of $J(\pi)$ between which there are no critical values of $J(\pi)$, the connectivities of the domains $(a, \rho) = A$ and $(b, \rho) = B$ are the same.*

* By an $i$-cycle ($i>0$) is meant a set of $i$-circuits. A 0-cycle shall mean any finite set of points.

† That is, homologies which are not empty (mod 2).

It will be sufficient to prove that a complete $j$-set for $A$, say $(K)_j$, is a complete $j$-set for $B$.

That every $j$-cycle on $B$ is homologous to a linear combination of $j$-cycles of $(K)_j$ follows from the fact that the domain $B$ can be $J$-deformed on itself into a set of points on $A$. It remains to prove that there are no homologies on $B$ between the members of $(K)_j$.

If now a $j$-cycle, say $H_j$, composed of $j$-cycles of $(K)_j$ bounded a complex $H_{j+1}$ on $B$, it would follow from Lemma 5.2 that we could $J$-deform $H_{j+1}$ on $B$ into a complex $H'_{j+1}$ on $A$. The boundary of $H'_{j+1}$ would be homologous on $A$ to $H_j$, so that $H_j$ would be bounding on $A$ contrary to the choice of $(K)_j$.

Thus $(K)_j$ is a complete $j$-set for $B$. The theorem follows at once.

7. **The preliminary homologies.** Let $a$ and $b$ be two non-critical values of $J(\pi)$ between which $J(\pi)$ takes on just one critical value $J(\pi)=c$. Let us suppose that this critical value corresponds to a single extremal $g$, of type $k$, joining $P$ to $Q$. We are going to show that the $J$-connectivities of the domains $A=(a,\rho)$ and $B=(b,\rho)$ differ only when $j=k$ or $k-1$.

*Let $\omega$ denote the set of critical points of $J(\pi)$ on $B$ at which $J(\pi)$ takes on the value $J(\pi)=c$.*

Each point $(\pi)$ of $\omega$ will determine the above extremal $g$. We come now to the following lemma.

LEMMA 7.1. *If there be given on $B$ an arbitrary neighborhood $R$ of the critical points $\omega$, then in any sub-neighborhood $R'$ sufficiently small every $j$-circuit is homologous to zero on $R$.*

It will be sufficient to show that, when suitably chosen, $R'$ can be continuously deformed on $R$ into a point on $R$. We shall prove that the required deformation can be effected by the product of three deformations $D_1D_3D_1$ of which $D_3$ will now be defined.

*The deformation $D_3$.* Let the extremal $g$ be slightly extended at its ends to form an extremal $g'$. From each of the $n$ vertices of points $(\pi)$ defining a broken extremal near $g$ we drop a perpendicular on $S$ to $g'$. We let the vertices of $(\pi)$ move along these perpendiculars towards $g'$ with velocities equal to the lengths of the perpendiculars. As the time $t$ varies from 0 to 1 the points $(\pi)$ with vertices near $g'$ will be deformed into points $(\pi)$ with vertices on $g'$. This is the deformation $D_3$.

Observe now that the deformation $D_1$ will carry each point of $\omega$ into that point $(\pi)_0$ of $\omega$ whose vertices divide $g$ into $n+1$ segments of equal $J$-length. Let $N$ be a neighborhood of the point $(\pi)_0$ so small that the deformation $D_3$ will carry points $(\pi)$ on $N$ into points $(\pi)$ defining $g$, rather than $g$ multiply covered in part.

If now $R'$ be sufficiently small, $D_1$, being continuous relative to different points $(\pi)$, will deform $R'$ on $R$ into points near $(\pi)_0$, or in particular into $N$. Under $D_3$ the points $(\pi)$ on $N$ will be deformed on $R$ into points $(\pi)$ defining $g$, while the final application of $D_1$ will carry the resultant points $(\pi)$ into the point $(\pi)_0$. Thus $D_1D_3D_1$ will serve as the required deformation.

In terms of the critical value $c$ we have the following lemma.

LEMMA 7.2. *If $R'$ be any neighborhood on $B-A$ of the points of the critical set $\omega$, then any $j$-circuit $K_j$ on $B$ satisfies an homology,*

$$K_j \sim C_j^1 + C_j^2, \text{ on } B,$$

*where $C_j^1$ is a complex on $R'$ and $C_j^2$ satisfies $J(\pi) < c - e_1$, where $e_1$ is a positive constant which depends only on $B$.*

To prove the theorem we shall first show that a sufficiently large number of iterations of the deformation $T$ of Lemma 5.2 will carry each point of $B$ either into $R'$, or else into a point at which $J(\pi)$ is less than $c$.

Note first that $T$ carries each point of $\omega$ into a point of $\omega$ and recall that $T$ is continuous. Let $R$ be a sub-neighborhood of $R'$ so small that $T$ will carry no point of $R$ outside of $R'$.

The set of points $B-A-R$ has a positive distance from the critical points of $J(\pi)$, and according to Lemma 5.2 will then be carried by $T$ into points $(\pi')$ at which

$$J(\pi') < J(\pi) - d, \qquad d > 0,$$

where it will be convenient to suppose $d < c - a$.

Now a sufficiently large number of iterations of $T$, say $T^r$, will carry $B$ into a set of points $(\pi)$ which satisfy

$$J(\pi) < c + \frac{d}{2}.$$

It follows that $T^rT$ will carry all points of $B$ whose $r$th images are not on $R$ into points at which

$$J(\pi) < c + \frac{d}{2} - d = c - \frac{d}{2} > a, \tag{7.1}$$

while the remaining points of $B$ will be carried into $R'$.

Let $K_j^1$ be the image of $K_j$ under $T^rT$. We now take $C_j^2$ as the sum of the $j$-cells on $K_j^1$ which satisfy (7.1) at some one point at least, and take for $C_j^1$ the sum of the remaining $j$-cells of $K_j^1$. Thus $C_j^1$ will lie on $R'$. If $K_j^1$ be sufficiently finely divided it appears from (7.1) that on $C_j^2$, $J(\pi)$ will be less than $c - d/4$. Thus the lemma is proved.

We state the obvious extension.

LEMMA 7.3. *If $R'$ be a neighborhood on $B-A$ of the points $\omega$, then any complex $K_j$ on $B$ whose boundary is on $A$, is homologous on $B$ to a complex $C_j^1+C_j^2$, which has the same boundary as $K_j$, and where $C_j^1$ and $C_j^2$ are complexes as in Lemma 7.2.*

8. **Deformations neighboring critical points.** Let us return to the space of the $m$ variables $(x)$. In an earlier paper we have proved for the plane essentially the following. (See Theorem 9 Morse II.*)

LEMMA 8.1. *On the extremal $g$ let there be $k$ points conjugate to $P$, with $k>0$, and $Q$ not conjugate to $A$.*

*Then corresponding to any sufficiently small neighborhood $N$ of $g$, there exists within $N$ an arbitrarily small neighborhood $N'$ of $g$, such that closed $m$-families of curves which join $g$'s end points on $N'$, and satisfy*

$$J \leqq c - e, \tag{8.1}$$

*where $e$ is a sufficiently small positive constant, are conditioned as follows:*

(a) *Those for which $m \neq k-1$ can be deformed on $N$, without increasing $J$, into a family of lower dimensionality.*

(b) *Those for which $m=k-1$ include a $(k-1)$-family $Z_{k-1}$ composed of admissible broken extremals, which is non-bounding† on $N$ and (8.1), and which is such that every other $(k-1)$-family can be deformed on $N$ without increasing $J$, either into $Z_{k-1}$ or else into a single curve.*

The proof of this theorem in the plane depended in the first instance upon Theorem 2 of Morse II. The later theorem has been generalized for $m$-space and appears as Theorem 2 of Morse III.

We note the following differences between the present theorem and Theorem 9 of Morse II.

The deformations here are affirmed not to increase $J$, instead of simply satisfying (8.1). The deformations used in Morse II are products of the deformation $D'$ of §29, Morse II, which as defined does not increase $J$, and of the deformation (3) of §31, Morse II. Now the latter deformation does increase $J$, but if in §31, (2) and (3), we replace $e^2$ by $a^2$, this deformation does not increase $J$, while the remainder of §31 holds as before. The remaining deformations of §31 are on $S_{k-1}$ of Morse II, and hence will not increase $J$.

---

* As far as the developments of the paper Morse II are concerned, and also for the present paper, we need not restrict the nature of $m$-families by requiring that the domain of the parametric point be a manifold, but simply that it be a complex, and this extension we suppose made.

† That is, whose "parametric complex" is non-bounding relative to other admissible parametric complexes.

Returning to the space of the points $(\pi)$ we have the following theorem.

THEOREM 2. *On the extremal $g$ suppose there are $k$ points conjugate to $P$, with $k>0$, and $P$ not conjugate to $Q$.*

*Then corresponding to any sufficiently small neighborhood $R$ of the critical set $\omega$, there exists within $R$ an arbitrarily small neighborhood $R'$ of $\omega$, such that the circuits $K_m$ on $R'$ which satisfy*

$$J(\pi) \leqq c - e, \tag{8.2}$$

*where $e$ is a sufficiently small positive constant, are conditioned as follows:*

(a) *Those for which $m \neq k-1$ are homologous to zero on $R$ and* (8.2).

(b) *Those for which $m=k-1$ include a $(k-1)$-circuit $D_{k-1}$ that is not homologous to zero on $R$ and* (8.2).

(c) *Those for which $m=k-1$ are either homologous to zero or to $D_{k-1}$, on $R$ and* (8.2).

This theorem will follow from Theorem 1 once we have examined the relations between neighborhoods $R$ of the set $\omega$ in the space $(\pi)$ and neighborhoods $N$ of $g$ in the space $(x)$.

Corresponding to a variable neighborhood $R$ of the set $\omega$ there exists a variable neighborhood $N$ of $g$ that contains all the broken extremals determined by points $(\pi)$ on $R$ and which approaches $g$ as $R$ approaches $\omega$.

Corresponding to a variable neighborhood $N$ of $g$ there exists a variable neighborhood $R$ of $\omega$ that contains all the points $(\pi)$ determined by those broken extremals on $N$ which satisfy $J<c$, and which approaches $\omega$ as $N$ approaches $g$. This follows from the generalization in $m$-space of (B) §29, Morse II.

With this understood the theorem follows readily.

## II. THE ANALYSIS SITUS

**9. Homologies among $j$-complexes when $j$ is not the type number.** For future use we now choose on the domain $B$ distinct from $A$ two closed neighborhoods $R$ and $R'$ of the critical points $\omega$ on $B$.

(a) *We choose $R$ so small that any circuit on $R$ is homologous to zero on $B$. See Lemma 7.1. We also choose $R$ so as to be admissible in Theorem 2.*

(b) *We choose $R'$ so small that any circuit on $R'$ is homologous to zero on some region interior to $R$.*

(c) *If $k \neq 0$ we still further restrict $R'$ so that $R$, $R'$, and a positive constant $e$, taken less than $c-a$, satisfy Theorem 2.*

(d) *If $k=0$ we restrict $R'$ so that on $R'$, $J(\pi)>c$ except at the critical points $\omega$ and their limit points.*

Let $C_j$ be any complex on $B$ whose boundary $C_{j-1}$ lies on $A$. An application of Lemma 7.3 gives the following homology and congruence:

$$C_j \sim C_j^1 + C_j^2 \quad \text{on } B, \tag{9.1}$$

$$C_{j-1} \equiv C_j^1 + C_j^2 \quad \text{on } B, \tag{9.2}$$

where $C_j^1$ is on $R'$ and $C_j^2$ satisfies

$$J(\pi) < c - e_1, \quad e_1 < e, \tag{9.3}$$

where $e_1$ is any sufficiently small positive constant. We have taken $e_1 < e$. Suppose $C_{j-1}^1$ is the common boundary* of $C_j^1$ and $C_j^2$. We have then

$$C_j^1 \equiv C_{j-1}^1 \quad \text{on } R', \tag{9.4}$$

$$C_j^2 \equiv C_{j-1}^1 + C_{j-1} \quad \text{on } B \text{ and } (9.3). \tag{9.5}$$

LEMMA 9.1. *If $j \neq k$, $C_j$ is homologous on $B$ to some $j$-complex on $A$.*

We divide the proof into three cases.

*Case* 1. $j \neq 0$, $k \neq 0$. Because $R$, $R'$, and $e$ satisfy Theorem 2, and because $j \neq k$, there exists a complex $C_j^3$ such that [see Theorem 2, (a)]

$$C_{j-1}^1 \equiv C_j^3 \quad \text{on } R \text{ and } (9.3). \tag{9.6}$$

From (9.4) and (9.6) we see that $C_j^1 + C_j^3$ is without boundary and on $R$. Because $R$ satisfies the restriction (a) we have then

$$C_j^1 + C_1^3 \sim 0 \quad \text{on } B. \tag{9.7}$$

From (9.1) and (9.7) we have

$$C_j \sim C_j^2 + C_j^3 \quad \text{on } B. \tag{9.8}$$

Now both $C_j^2$ and $C_j^3$ satisfy (9.3). With the aid of $J$-deformations we see that $C_j^2 + C_j^3$ is homologous to a complex $K_j$ on $A$, with the same boundary on $A$. Hence

$$C_j \sim K_j \quad \text{on } B$$

and the proof is complete in Case 1.

*Case* 2. $j = 0$, $k \neq 0$. For $j = 0$, (9.1) still holds. We next define $C_j^3$ as any set of points on $R$ and on (9.3) equal in number to the points of $C_j^1$. The proof may now be started with (9.7), and proceeds therefrom as before.

*Case* 3. $j \neq 0$, $k = 0$. Here $C_{j-1}^1$ lies on $R'$ and satisfies (9.3), which offers a contradiction to (d) unless $C_{j-1}^1$ is null. Thus $C_{j-1}^1$ is null. We now proceed exactly as in Case 1, understanding, however, that both members of (9.6) are now null.

* We admit the possibility that $C_{j-1}$ or $C_{j-1}^1$ be null. For $j = 0$ we regard them as meaningless.

10. **Homologies among $k$-circuits when $k$ is the type number ($k \neq 0$).** We understand that $R$ and $R'$ are chosen as in the preceding section. According to Theorem 2 there then exists a $(k-1)$-cycle $D_{k-1}$ on $R'$ and (9.3) which bounds no complex on $R$ and (9.3). Since $R'$ satisfies (b) of §9 we have, however,

$$D_{k-1} \equiv D_k^1 \quad \text{on } R, \tag{10.1}$$

where $D_k^1$ is a $k$-complex interior to $R$.

We now come to a division into two cases according to the nature of $B$.

*Case $\alpha$. There exists a complex $D_k^2$ such that*

$$D_{k-1} \equiv D_k^2 \quad \textit{on } (9.3) \textit{ and } B. \tag{10.2}$$

*Case $\beta$. No such complex as $D_k^2$ exists.*

LEMMA 10.1. *In Case $\alpha$, the $k$-cycle $D_k$ defined as*

$$D_k^1 + D_k^2 = D_k \quad \textit{on } B \tag{10.3}$$

*is not homologous on $B$ to any cycle* on $A$.*

Suppose the lemma false and that $D_k$ were homologous on $B$ to a $k$-cycle $D_k^3$ lying on $A$. We would have then

$$D_k^1 + D_k^2 + D_k^3 \equiv D_{k+1} \quad \text{on } B \tag{10.4}$$

where $D_{k+1}$ is a complex on $B$.

We can further suppose that $D_{k+1}$ consists of points which either satisfy (9.3), or else are interior to $R$. For as in §7 we could $J$-deform $D_{k+1}$ into such a complex. Such a deformation, say $T'$, would alter the boundary of $D_{k+1}$, but it could be replaced by a deformation which did not alter this boundary. One would need simply to hold the boundary fast during $T'$ and hold points near this boundary fast during a portion of the deformation starting from a suitable time depending continuously on the distance to the boundary.

With this understood, let $D_k^4$ be the boundary of the sum of those $(k+1)$-cells of $D_{k+1}$ which do not wholly satisfy (9.3), or else whose boundaries do not wholly satisfy (9.3). We see that $D_k^4$ is on (9.3) except for such $k$-cells as it has in common with $D_k^1$. If $D_{k+1}$ has been sufficiently finely divided $D_k^4$ would also lie on $R$, as we now suppose. Thus

$$D_k^1 + D_k^4 \tag{10.5}$$

is on $R$, and if reduced, mod 2, satisfies (9.3). But since $D_k^4$ is without boundary, we have from (10.1)

* Including the null cycle.

$$D_{k-1} \equiv D_k^1 + D_k^4. \tag{10.6}$$

Thus $D_{k-1}$ would bound the complex (10.5) which is on $R$ and (9.3), contrary to the choice of $D_{k-1}$.

The lemma is thereby proved.

To proceed further it will be convenient to divide the complexes $C_k$ of §9 ($k \neq 0$) into two classes according to the nature of the complex $C_{k-1}^1$ associated with $C_k$ by (9.4) and (9.5). The distinctive properties of these classes are the following:

Class I. $C_{k-1}^1 \sim 0$ on $R$ and (9.3).

Class II. $C_{k-1}^1 \sim D_{k-1}$ on $R$ and (9.3).

That this division is exhaustive follows from Theorem 2.

LEMMA 10.2. *In Case $\alpha$, each $k$-cycle $C_k$ on $B$ is homologous on $B$ to a linear combination of the cycle $D_k$ and cycles on $A$.*

*$C_k$ in Class* I. Here there exists a complex $C_k^3$ such that

$$C_{k-1}^1 \equiv C_k^3 \quad \text{on } R \text{ and } (9.3). \tag{10.7}$$

According to (9.4) and (10.7), $C_k^1 + C_k^3$ is a $k$-cycle. Since it is on $R$ we have

$$C_k^1 + C_k^3 \sim 0 \quad \text{on } B. \tag{10.8}$$

From (9.1) and (10.8) we have respectively

$$C_k \sim C_k^1 + C_k^2 \sim C_k^3 + C_k^2 \quad \text{on } B. \tag{10.9}$$

Now $C_k^3 + C_k^2$ satisfies (9.3), and so can be $J$-deformed on $B$ into a $k$-cycle $H_k$ on $A$. We thus have

$$C_k \sim H_k \quad \text{on } B.$$

Thus the lemma is proved when $C_k$ is in Class I.

*$C_k$ in Class* II. Here we have

$$C_{k-1}^1 + D_{k-1} \equiv C_k^4 \quad \text{on } R \text{ and } (9.3), \tag{10.10}$$

where $C_k^4$ is a complex on $R$ and (9.3). Adding (9.4) for $k=j$, (10.1), and (10.10), we have

$$D_k^1 + C_k^1 + C_k^4 \equiv 0 \quad \text{on } R. \tag{10.11}$$

Now the complex (10.11) is a $k$-cycle on $R$ so that

$$D_k^1 + C_k^1 + C_k^4 \sim 0 \quad \text{on } B. \tag{10.12}$$

From the definition of $D_k$ and from (9.1) we have

$$D_k + C_k \sim D_k^1 + D_k^2 + C_k^1 + C_k^2 \quad \text{on } B. \tag{10.13}$$

Eliminating $C_k^1$ from (10.13) with the aid of (10.12) we have

$$D_k + C_k \sim D_k^2 + C_k^4 + C_k^2 \quad \text{on } B. \tag{10.14}$$

But the right member of (10.14) satisfies (9.3) and is accordingly homologous on $B$ to a cycle $H_k$ on $A$. With the aid of (10.14) we have then

$$C_k \sim D_k + H_k \quad \text{on } B,$$

and the lemma is proved when $C_k$ is in Class II.

LEMMA 10.3. *In Case $\beta$ each $k$-cycle $C_k$ on $B$ is homologous on $B$ to a $k$-cycle on $A$.*

If $C_k$ belongs to Class I the proof in the preceding lemma applies. We shall show that in Case $\beta$, Class II is empty.

If $C_k$ belonged to Class II, (10.10) would hold as described. Adding (10.10) and (9.5) for $j=k$, and noting that $C_{k-1}=0$, we have

$$D_{k-1} \equiv C_k^2 + C_k^4 \quad \text{on } B \text{ and } (9.3),$$

contrary to the hypothesis distinguishing Case $\beta$.

11. **Homologies among $(k-1)$-circuits, $k\neq 0$.** In this section we prove three lemmas.

LEMMA 11. 1. *In Case $\alpha$ any cycle $C_{k-1}$ on $A$ which is homologous to zero on $B$ will be homologous to zero on $A$.*

If $C_{k-1}$ is homologous to zero on $B$ it bounds a complex $C_k$ which we identify with the complex $C_j$ of §9. We turn next to $C_k^1{}_{-1}$ defined by (9.4) and (9.5). According to Theorem 2, $C_k^1{}_{-1}$ is either homologous to zero or to $D_{k-1}$, on $R$ and (9.3). If $C_k^1{}_{-1}$ is homologus to $D_{k-1}$ we note that, in Case $\alpha$, $D_{k-1}$ is homologous to zero on $B$ and (9.3). It is always true then in Case $\alpha$ that there exists a $k$-complex $C_k^5$ such that

$$C_k^1{}_{-1} \equiv C_k^5 \quad \text{on } B \text{ and } (9.3). \tag{11.1}$$

Upon adding (9.2), (9.4) and (11.1) we have

$$C_{k-1} \equiv C_k^2 + C_k^5 \quad \text{on } B \text{ and } (9.3). \tag{11.2}$$

Since $C_k^2 + C_k^5$ satisfies (9.3) it can be $J$-deformed into a complex $K_k$ on $A$ without altering $C_{k-1}$. Thus $C_{k-1}$ bounds $K_k$ on $A$ and the lemma is proved.

*Definition of $D_k^1{}_{-1}$.* Denote by $D_k^1{}_{-1}$ a $(k-1)$-cycle obtained by deforming $D_{k-1}$ on $B$ and (9.3) into a $(k-1)$-circuit on $A$. Such a deformation is possible since $D_k^1{}_{-1}$ satisfies (9.3). According to the definition of $D_k^1{}_{-1}$ we have

$$D_k^1{}_{-1} + D_{k-1} \equiv D_k^5 \quad \text{on } B \text{ and } (9.3), \tag{11.3}$$

where $D_k^5$ is a $k$-complex on $B$ and (9.3). We note that

$$D_{k-1}^1 \sim 0 \quad \text{on } B. \tag{11.4}$$

For upon adding (11.3) and (10.1) we have

$$D_{k-1}^1 \equiv D_k^1 + D_k^5 \quad \text{on } B.$$

LEMMA 11.2. *In Case $\alpha$, $D_{k-1}^1$ is homologous to zero on $A$, but in Case $\beta$, $D_{k-1}^1$ is not homologous to zero on $A$.*

That $D_{k-1}^1$ is homologous to zero on $A$ in Case $\alpha$ follows from (11.4) and Lemma 11.1.

If in Case $\beta$ we had a congruence

$$D_{k-1}^1 \equiv D_k^6, \tag{11.5}$$

where $D_k^6$ is a $k$-complex on $A$, then upon adding (11.3) and (11.5) we would have

$$D_{k-1} \equiv D_k^5 + D_k^6.$$

But $D_k^5 + D_k^6$ satisfies (9.3) contrary to the nature of $D_{k-1}$ in Case $\beta$, and the lemma is proved.

LEMMA 11.3. *In Case $\beta$, any cycle $C_{k-1}$ on $A$ which is homologous to zero on $B$ is either homologous to zero or to $D_{k-1}^1$ on $A$.*

If $C_{k-1}$ is homologous to zero on $B$ it bounds a complex $C_k$ which can be identified with the complex $C_j$ of §9.

If $C_k$ is of Class I we repeat the reasoning of Lemma 10.2 under Class I except that we here follow (10.9) with the statement that $C_k^3 + C_k^2$ can be $J$-deformed on $B$ into a complex $K_k$ on $A$ without altering $C_{k-1}$. Thus $C_{k-1}$ bounds $K_k$ on $A$ and the lemma is proved if $C_k$ is in Class I.

If $C_k$ is in Class II, (10.10) holds as before. Upon adding (9.2) and (9.4), for $j = k$, to (10.10) and (11.3), we have

$$D_{k-1}^1 + C_{k-1} \equiv D_k^5 + C_k^4 + C_k^2. \tag{11.6}$$

But the right member of (11.6) satisfies (9.3), from which fact we infer that

$$D_{k-1}^1 \sim C_{k-1} \quad \text{on } A$$

and the proof is complete.

12. **The changes in the connectivities when $k \neq 0$.** In this section we prove two theorems.

THEOREM 3. *Let $a$ and $b$ be two non-critical values of $J(\pi)$, $a<b$, between which $J$ takes on a critical value corresponding to just one extremal of type $k$. The connectivities $R_j$ of the domains $A=(a, \rho)$ and $B=(b, \rho)$ differ at most when $j=k$ or $k-1$.*

If $(K)_j$ be a complete $j$-set for $A$, it follows from Lemma 9.1 for $j\neq k$ that any $j$-cycle on $B$ is homologous on $B$ to a linear combination of members of $(K)_j$. From the same lemma it follows for $j\neq k-1$ that there can be no proper homologies on $B$ among the members of $(K)_j$ without there being proper homologies among the members of $(K)_j$ on $A$. Thus $(K)_j$ is a complete $j$-set for $B$ and the theorem is proved.

THEOREM 4. *If $k\neq 0$ the connectivities $R_j$ of the domains $A$ and $B$ differ only as follows*:

(12.1) *Case* $\alpha$ : $\Delta R_k = 1$;

(12.2) *Case* $\beta$ : $\Delta R_{k-1} = -1$,

*where $\Delta$ refers to the change as we pass from $A$ to $B$.*

*We shall first treat Case $\alpha$.*

Let $(K)_k$ be a complete $k$-set for $A$. I say that $(K)_k$ together with $D_k$ of §10 will form a complete $k$-set for $B$.

For it follows first from Lemma 10.2 that every $k$-cycle on $B$ is homologous on $B$ to a linear combination of $D_k$ and the members of $(K)_k$.

Secondly there can be on $B$ no proper homologies involving the members of $(K)_k$ and $D_k$. For if a combination of members of $(K)_k$ alone were bounding on $B$, with the aid of Lemma 9.1 we could infer that the same combination would be bounding on $A$, contrary to the nature of $(K)_k$. But no linear combination of $D_k$ and the members of $(K)_k$ which actually involved $D_k$ could be homologous to zero on $B$ without contradicting Lemma 10.1.

Thus $(K)_k$ with $D_k$ forms a complete $k$-set for $B$ so that $\Delta R_k=1$.

We must now show that $\Delta R_j=0$ when $j\neq k$. This is already established in Theorem 3 provided $j\neq k-1$. When $j=k-1$, Lemmas 9.1 and 11.1 show that a complete $j$-set for $A$ is a complete $j$-set for $B$. The theorem is thus established for Case $\alpha$.

*We can now treat Case $\beta$.*

Turning to Lemma 10.3 and Theorem 3 we see readily that $\Delta R_j=0$ for $j\neq k-1$.

We shall show finally that (12.2) holds.

According to Lemma 11.2, $D_{k-1}^1$ is not homologous to zero on $A$, and so there exists a complete $(k-1)$-set for $A$, say $(K)_{k-1}$, of which $D_{k-1}^1$ is a mem-

ber. Let $(H)_{k-1}$ be the set of $(k-1)$-circuits in this set after $D^1_{k-1}$ has been removed. I say that $(H)_{k-1}$ is a complete $(k-1)$-set for $B$.

For according to Lemma 9.1 and the homology (11.4) any $(k-1)$-cycle on $B$ is homologous on $B$ to a combination of $(k-1)$-cycles from $(H)_{k-1}$. That there are no homologies on $B$ among members of $(H)_{k-1}$ follows readily from Lemma 11.3. Thus (12.2) holds as required and the theorem is proved.

13. **The case of minima, $k=0$.** Here the value $J(\pi)=c$ furnishes a relative minimum to $J(\pi)$ on $B$. We understand a relative minimum to include an absolute minimum as a special case.

*Let $P$ be a point of the critical set $\omega$.*

LEMMA 13.1. *If $c$ is a relative minimum for $J(\pi)$ on $B$, then any 0-cycle $C_0$ on $B$ is homologous on $B$ to a 0-cycle on $A$, or a combination of 0-cycles on $A$ and the 0-cycle $P$.*

According to Lemma 7.3 we have the homology

$$C_0 \sim C_0^1 + C_0^2 \quad \text{on } B, \tag{13.1}$$

where $C_0^1$ is a set of $r$ points on $R'$, and $C_0^2$, possibly null, is a set of points on $B$ at which $J(\pi)<c$. Note first that $C_0^2$ is homologous on $B$ to a set $K_0$ of points on $A$ so that

$$C_0 \sim C_0^1 + K_0 \quad \text{on } B. \tag{13.2}$$

We distinguish between the cases $r$ even and $r$ odd.

If $r$ is even, $C_0^1$, according to the choice of $R'$, is homologous to zero on $B$, so that $C_0$ is homologous to $K_0$ on $B$ and the lemma is proved. If $r$ is odd, we have

$$C_0 \sim P + K_0 \quad \text{on } B,$$

and the lemma is proved for all cases.

LEMMA 13.2. *If $J(\pi)=c$ is a relative minimum for $J(\pi)$ on $B$, the point $P$ on $\omega$ cannot be connected on $B$ with any point on $A$.*

To prove this, note that at the boundary points of $R'$ which are not also boundary points of $B$, $J(\pi)>c+e$, where $e$ is a positive constant.

Suppose $P$ could be joined to a point on $A$ by a continuous curve $C_1$ on $B$. A sufficient number of iterations of the deformation $T$ would carry $C_1$ into a curve $C_1^1$ on $B$ on which $J(\pi)<c+e$. But it would also carry $P$ into a point on $\omega$ in $R'$, so that $C_1^1$ would still join a point on $\omega$ in $R'$ to a point on $A$. This is impossible, since on $C_1^1$, $J(\pi)<c+e$. Thus the lemma is proved.

If $J(\pi)=c$ gives an absolute minimum to $J(\pi)$ on $B$, $A$ will be devoid of points. In that case we say that the *connectivities of $A$ are all zero.* With this understood we have the following theorem.

THEOREM 5. *If $k=0$ the connectivities $R_i$ of the domains $A$ and $B$ differ only in that $\Delta R_0=1$, where $\Delta R_0$ is the change in $R_0$ as we pass from $A$ to $B$.*

That $R_0$ is the only connectivity to change as we pass from $A$ to $B$ follows from Theorem 3.

Now if $(K)_0$ is a complete 0-set for $A$, null if $A$ is null, then $(K)_0$ with $P$ will be a complete 0-set $(H)_0$ for $B$.

That all 0-cycles on $B$ are homologous on $B$ to members of $(H)_0$ is affirmed by Lemma 13.1.

That there are no homologies on $B$ among members of $(H)_0$ can be seen as follows. There are no homologies on $B$ among members of $(K)_0$. For such homologies would lead, upon applying Lemma 9.1 for $j=1$, to the conclusion that there were homologies on $A$ among the same members of $(K)_0$. Any homology on $B$ among the members of $(H)_0$ must then explicitly involve $P$ which would mean that $P$ on $\omega$ would be connected on $B$ to some point on $A$ contrary to Lemma 13.2. Thus there are no homologies on $B$ among the members of $(H)_0$.

Thus $(K)_0$ and $P$ form a complete 0-set for $B$. The theorem follows at once.

## III. THE EXISTENCE OF EXTREMALS AND THEIR RELATIONS

14. **Relations between the extremals and the connectivities.** *It will be convenient to call the extremal appearing in Theorem 4 of increasing or decreasing type according as $\Delta R_k=1$ or $\Delta R_{k-1}=-1$.*

In accordance with Theorem 5 each extremal of type zero should be classified as increasing in type.

We come now to the following principal theorems. As previously we suppose $a$ is a non-critical value of $J$, and that $P\neq Q$.

THEOREM 6. *If among the extremals $g$ on which $J<a$ there are none on which $P$ is conjugate to $Q$, then between the integers $M_k$ which give the total number of these extremals of type $k$, and the connectivities $R_i$ of the domain $(a,\rho)$, the following relations hold:*

$$
\begin{aligned}
M_0 &\geqq R_0,\\
M_0 - M_1 &\leqq R_0 - R_1,\\
M_0 - M_1 + M_2 &\geqq R_0 - R_1 + R_2, \qquad (14.1)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
M_0 - M_1 + \cdots + (-1)^r M_r &= R_0 - R_1 + \cdots + (-1)^r R_r,
\end{aligned}
$$

*where $r$ is the maximum of the type numbers $k$.*

Consider first the case where the $J$-lengths of the different extremals $g$ are all different.

Let the extremals of type $i$ be further divided into $p_i$ extremals of increasing type, and $q_i$ extremals of decreasing type. We have at once

$$M_i = p_i + q_i, \tag{14.2}$$

$$R_i = p_i - q_{i+1} \qquad (i = 0, 1, \cdots, r), \tag{14.3}$$

where, in particular, $q_0$ and $q_{r+1}$ equal zero. From the relations (14.2) and (14.3) we may eliminate $p_i$ and find that

$$M_0 - M_1 + \cdots + (-1)^i M_i = R_0 - R_1 + \cdots + (-1)^i R_i + (-1)^i q_{i+1}. \tag{14.4}$$

Relations (14.1) follow at once from (14.4).

If now two or more of the extremals joining $P$ to $Q$ have the same $J$-length it follows from Lemma 16.2 that $Q$ can be replaced by a point $Q'$, arbitrarily near $Q$, and such that for $Q'$ the $J$-lengths of the different extremals $g$ are all different.

For the pair $PQ'$ the relations (14.1) now hold. But as $Q'$ approaches and takes the position $Q$, the numbers $M_k$ will not change, since $P$ is not conjugate to $Q$ on any of the extremals $g$. Neither will the connectivities $R_i$ change, as follows from Theorem 13 of §18.

Thus the relations (14.1) must hold in all cases and the theorem is proved.

We have the following corollary.

COROLLARY. *If the ith connectivity of $(a, \rho)$ is $R_i$ there exist at least $R_i$ extremals joining $P$ to $Q$ of type $i$.*

To proceed further it will be convenient to introduce the conception of the *conjugate sequence* for the pair $PQ$.

Let $P$ and $Q$ be any pair of points on $S$. Let $N_k$ be the number of closed extremal segments joining $P$ to $Q$ on which there are $k$ points conjugate to $P$, counting conjugate points according to their orders. If there are an infinity of such extremals we replace $N_k$ by $\infty$.

*The sequence*

$$N_0 N_1 N_2 \cdots \tag{14.5}$$

*will be called the conjugate sequence for $P$ and $Q$.*

We shall investigate this sequence first for pairs $P$ and $Q$ which are non-specialized in the following sense.

*Two distinct points $P$ and $Q$ which are not conjugate to each other on any extremals will be termed non-specialized.*

In §16 we shall prove the following.

*There exist non-specialized pairs of points PQ in the respective neighborhoods of any two points on S.*

We shall begin by supposing $S$ restricted to the most important particular case.

15. **The case where the region $S$ is elementary.** We suppose now that $S$ is an elementary region in the sense that it is homeomorphic with an $m$-sphere and its interior.

For the moment let us term a curve on $s$ *admissible* if it has a continuously turning tangent except at most at a finite number of points, and joins $P$ to $Q$. The fact that $S$ is elementary as well as extremal-convex leads to the following lemma.

LEMMA 15.1. *On S there exists a continuous deformation $\delta$ that deforms all admissible curves for which $J<a$ through admissible curves of J-lengths less than some constant b, into a single admissible curve, thereby deforming continuous families of such curves continuously.*

Since $S$ is elementary there exists a continuous deformation, say $\delta_1$, of $S$ on itself, leaving $P$ and $Q$ fixed, that carries $S$ into a set of points on a curve $\gamma$ joining $P$ to $Q$. We can moreover suppose, for simplicity, that $\gamma$ is a minimizing extremal joining $P$ to $Q$.

Now let each admissible curve $g$ for which $J<a$ be divided into $r+1$ successive segments $g_i$ of equal $J$-length. Let us suppose $r$ taken so large that the end points of $g_i$ and their images under $\delta_1$ can be joined by elementary extremals.

We now define the deformation $\delta$. On each curve $g$ let each segment $g_i$ be replaced by the elementary extremal $g_i'$ that joins its end points, or more particularly let a point on $g_i$ dividing $g_i$ in a certain ratio be replaced by that point on $g_i'$ which divides $g_i'$ in the same ratio. Each curve $g$ can be readily deformed through admissible curves into the corresponding curve $g'$ by a deformation that deforms continuous families of such curves continuously.* To deform further the resultant curves $g'$ into a single admissible curve we require the end points of $g_i'$ to move according to the deformation $\delta_1$, requiring a point on $g_i'$ which divides $g_i'$ in a certain ratio to be deformed into that point of the moving elementary extremal $g_i''$ which divides $g_i''$ in the same ratio.

Each curve $g$ will thereby be deformed into a broken extremal $g''$ with its vertices on $\gamma$. Finally let these vertices move along $\gamma$ each at a constant $J$-rate into a set of vertices which divide $\gamma$ into a set of elementary extremals

* See the deformation $D'$; Morse II p. 263.

of equal $J$-length, thereby deforming $g''$ into a single admissible curve $\gamma$ in the desired manner.

If the elementary extremals used have $J$-lengths at most $\rho$, the curves used in the deformation will have $J$-lengths at most $r\rho$. Thus $b$ can be any constant $b > \rho r$.

LEMMA 15.2. *The constant $b$ of Lemma* 15.1 *can be chosen independently of the position of $P$ and $Q$ on $S$.*

To prove this let $P'$ and $Q'$ be any other pair of points on $S$ and let $g$ be an admissible curve joining them. Let $p^+$ and $q^+$ be minimizing extremals joining $PP'$ and $QQ'$ respectively in the senses indicated. Holding $P'$ and $Q'$ fast let $g$ be deformed into the curve $g_1$ consisting of the sequence of curves $(p^- p^+ g q^- q^+)$ joining $P'$ to $Q'$. Note that this deformation can be made without increasing $J$-lengths by more than $4d$, where $d$ is the maximum of the $J$-lengths of minimizing extremals joining any two points of $S$.

Now the curve $(p^+ g q^-)$ gives a subsegment of $g_1$ joining $P$ to $Q$. The class of such subsegments can be deformed through admissible curves joining $P$ to $Q$ into a single curve $\gamma$ joining $P$ to $Q$, using thereby curves all of $J$-lengths less than some constant $b_1$, as is affirmed by the preceding lemma.

The class of curves $g$ will thereby be deformed into the curve $(p^- \gamma q^+)$ through the mediation of admissible curves joining $P'$ to $Q'$ all of $J$-lengths less than $b_1 + 4d$. The latter constant can be taken as our choice of $b$.

Thus the statement in italics is proved.

LEMMA 15.3. *All circuits on the domain $(a, \rho)$ are homologous to zero on a domain $(b, \rho)$ where $b$ is any constant sufficiently large chosen independently of $P$ and $Q$.*

It will be sufficient to choose $b$ as any constant at least as great as the constant $b$ of Lemma 15.2. As previously we suppose $J_0$ greater than $a$ and $b$, and $n$ chosen as in §3.

Let $C_i$ be any $i$-circuit on $(a, \rho)$. If the broken extremals corresponding to $C_i$ be subjected to the deformation $\delta$ of the preceding lemmas they will be carried into the curve $\gamma$. We can set up a corresponding deformation of the points $(\pi)$ of $C_i$ as follows.

We first apply the deformation $D$ of §5 to $C_i$. The end points of the resulting elementary extremals will divide each original curve $g$ defined by by $C_i$ into $n+1$ segments of equal $J$-lengths. Let these end points now take those positions on the variable curve $g'$ replacing $g$ during $\delta$ which divide $g'$ into segments of equal $J$-length. The resulting deformation will carry $C_i$ into a point $(\pi)_0$ corresponding to $\gamma$.

The lemma follows directly.

We come now to the following theorem.

THEOREM 7. *Corresponding to the set of extremals which join two non-specialized points $P$ and $Q$ on which $J<a$, there always exists a complementary set of extremals on which $a<J<b$ such that for the combined sets*

$$
\begin{aligned}
&m_0 \geqq 1,\\
&m_0 - m_1 \leqq 1,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&m_0 - m_1 + \cdots + (-1)^r m_r = 1 + (-1)^r R_r,
\end{aligned}
\tag{15.1}
$$

*where $m_i$ is the number of extremals of type $i$ in the combined sets and $r$ is the maximum of the type numbers of the extremals of the original set and where $R_r$ is the $r$th connectivity of $(a, \rho)$.*

As previously we need prove the theorem only for the case where the $J$-lengths of the different extremals joining $P$ to $Q$ are different.

The relations (14.1) of Theorem 6 furnish our starting point.

From Lemma 15.3 we see that all circuits on $(a, \rho)$ are homologous to zero on a domain $(b, \rho)$ for which $b$ is large enough.

If $R_i$ is the $i$th connectivity of $(a, \rho)$ we see that there must be at least $R_i$ extremals for $i>0$, and $R_0-1$ for $i=0$, of decreasing type $i+1$ with $a<J<b$. We take these extremals $(i=0, \cdots, r)$ as the complementary set. In terms of the integers $M_i$ and $R_i$ of Theorem 6 we have

$$
m_0 = M_0, \quad m_1 = M_1 + (R_0 - 1),
$$
$$
m_i = M_i + R_{i-1} \qquad (i = 1, \cdots, r).
$$

Relations (15.1) follow now with the aid of (14.1).

We note that, in Theorem 7, $b$ is a sufficiently large positive constant which may, in particular, be chosen independently of the position of $P$ and $Q$ on $S$.

We turn now to the conjugate sequences (14.5).

THEOREM 8. *If none of the integers $N_i$ of the conjugate sequence of a non-specialized pair of points are infinite, they satisfy the infinite set of inequalities*

$$
\begin{aligned}
&N_0 \geqq 1,\\
&N_0 - N_1 \leqq 1,\\
&N_0 - N_1 + N_2 \geqq 1,\\
&\cdots\cdots\cdots\cdots
\end{aligned}
\tag{15.2}
$$

*If all of the integers of the conjugate sequence are finite up to $N_{k+1}$, the first $k+1$ relations in (15.2) still hold.*

The first statement in the theorem is a consequence of the last. Let us turn then to the last.

Suppose all of the integers up to $N_{k+1}$ are finite. Let $a$ be a non-critical value of $J$ greater than the $J$-lengths of all extremals with types at most $k$. If now we apply Theorem 7 we find that we must have

$$m_i = N_i \qquad (i = 0, 1, \cdots, k)$$

and the first $k+1$ relations in (15.2) follow from (15.1).

The theorem follows directly.

COROLLARY. *If there are no extremals joining $P$ and $Q$ upon which there are $k$ conjugate points, then there are either an infinite number of extremals upon which there are fewer than $k$ conjugate points, or else the numbers $N_0, \cdots, N_{k-1}$ satisfy*

$$N_0 - N_1 + \cdots + (-1)^{k-1} N_{k-1} = 1.$$

This follows from the $k$th and $(k+1)$st inequalities of (15.2).

*We come now to a theorem in which we shall not require the numbers $N_i$ to be finite.*

If $N_r$ represents $\infty$ we shall understand $N_i - 1$ and $N_i + 1$ as also representing $\infty$.

THEOREM 9. *Let $P$ and $Q$ be any two non-specialized points on $S$.*

(a) *If there are $N_k$ extremals of type $k > 1$ there are at least $N_k$ extremals of the two adjacent types.*

(b) *If there are $N_1$ extremals of type one there are at least $N_1 - 1$ extremals of the two adjacent types.*

(c) *If there are $N_0$ extremals of type zero there are at least* $N_0 - 1$ extremals of the adjacent type one.*

We shall first prove (a).

From the first inequality in (15.1) which involves $m_{i+1}$, and the third preceding inequality, we find that

$$m_i \leqq m_{i-1} + m_{i+1}, \quad i > 1.$$

If $N_{i-1}$, $N_i$, and $N_{i+1}$ are finite, and we take the constant $a$ in Theorem 7 large enough, these $N$'s become equal to $m_{i-1}$, $m_i$ and $m_{i+1}$ respectively, and (a) is proved for this case.

* Part (c) is the representation here of the "minimax principle" of Birkhoff. See *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), p. 249.

If $N_i$ is infinite there must be extremals of type $i$ of arbitrarily great $J$-length. If we take $a$ successively as the constants of a sequence of constants becoming infinite, $m_i$ will become infinite, and hence from (15.1) either $m_{i-1}$ or else $m_{i+1}$. Part (a) then follows in this case as well.

Parts (c) and (b) follow similarly from the second and third inequalities in (15.1).

## IV. The density of conjugate points, and invariance of the connectivities

**16. Specialized points $P$ and $Q$.** We wish to show that the so called non-specialized pairs of points are really general. We begin with the following lemma.

LEMMA 16.1. *The conjugate points of a point $P$ at distances from $P$ along the corresponding extremals not exceeding a positive constant $d$, form a set which is nowhere dense.*

Let $(u)$ represent a point in an auxiliary $m$-space. Let each extremal issuing from $P$ be represented in the space $(u)$ by a ray issuing from the origin with a direction parallel to its direction at $P$, and with points on the extremal at distances $s$ from $P$, corresponding to points on the ray at distances $s$ from the origin. The corresponding functions

$$x_i = x_i(u) \qquad (i = 1, 2, \cdots, m)$$

will be analytic except at the origin. The jacobian $D(u)$ of these functions will vanish at the conjugate points. Its rank $r$ at such points will be between 0 and $m$ (Morse III §7).

Suppose $(x)_0$ is a conjugate point corresponding to a point $(u)_0$. If $D(u_0)$ is of rank $r$ one sees that there is an $r$-plane $X$ through $(x)_0$ in the space $(x)$, such that the distance of the points $[x(u)]$ from $X$ is an infinitesimal of at least the second order with respect to the distance $\rho$ of $(u)$ from $(u)_0$.

Let $S_1$ be the interior of an $(m-1)$-sphere of radius $\rho$ with center at $(u)_0$. One sees that the points $(x)$ corresponding to points $(u)$ on $S_1$ can be enclosed in a volume $V$ whose ratio to that of $S_1$ will approach zero as $\rho$ approaches zero. For $V$ one could take a generalized cylinder consisting of the points $P$ of $X$ at a distance $c\rho$ from $(x)_0$ together with the points on perpendiculars to $X$, points at a distance $h\rho^2$ from these points $P$, where $c$ and $h$ are suitably chosen positive constants independent of the choice of $(u)_0$.

Let $e$ now be an arbitrarily small positive constant. Let us break the space $(u)$ up into congruent $m$-cubes. If the diameter of each of these $m$-cubes be sufficiently small, then such of the corresponding sets $[x(u)]$ as contain conjugate points with $s \leqq d$ can be enclosed in elementary volumes

such as $V$ whose ratios to that of the cubes will be less than $e$. The sum of these volumes $V$ will be less than $e$ times the total volume of the corresponding cubes. The sum of these elementary volumes will then be arbitrarily small. This is possible only if the conjugate points in question are nowhere dense.

THEOREM 10. *The set of all conjugate points of a fixed point $P$ is nowhere dense on $S$.*

Let $d_1, d_2, \cdots$ be an increasing set of positive constants which become infinite with their subscripts. Let $Q$ be any point on $S$. It follows from the preceding lemma that there exists in any neighborhood of $Q$ a sequence of $(m-1)$-spheres

$$S_1, S_2, \cdots$$

each within the other, and such that there are within $S_i$ no points conjugate to $P$ for which $s<d_i$. These $(m-1)$-spheres will have at least one common interior point, say $A$. The point $A$ cannot be a conjugate point of $P$ without violating the principle under which the $(m-1)$-spheres $S_i$ were stated to exist.

Thus the theorem is proved.

This theorem amounts to the statement already made that there exist non-specialized pairs of points $P$ and $Q$ in the respective neighborhoods of any two given points of $S$.

The following theorem is a strong aid in proving the existence of extremals joining specialized pairs of points.

THEOREM 11. *If each pair of points in a set of non-specialized points can be joined by an extremal $\gamma$ which is bounded in $J$-length for the set, and of type $k$, then any limit pair $P_0Q_0, P_0 \neq Q_0$, of pairs of the set can be joined by an extremal $g$ on which there are at least $k$, and at most $k+m-1$, points* conjugate to $P_0$.*

Let $PQ$ represent any pair of points in the set. As $PQ$ approaches $P_0Q_0$, the initial directions of the extremals $\gamma$ will have at least one limit direction. Let $g$ be the extremal with this limit direction.

The extremal $g$ will have at least $k$ conjugate points on it. For otherwise there would be fewer than $k$ conjugate points on extremals $\gamma$ neighboring $g$. See Morse IV §9.

Suppose there were $k+r+s$ conjugate points on $g$, where $s$ is the number of conjugate points to be counted at $Q_0$. Now $r \leqq 0$, for otherwise extremals $\gamma$ neighboring $g$ would have at least $k+r>k$ conjugate points on them. Finally

* Counting conjugate points according to their orders.

$s \leqq m-1$ as was shown in Morse III §7. Thus there are at most $k+m-1$ conjugate points on $g$.

The following two lemmas have already been used. They can be conveniently proved here.

LEMMA 16.2. *Let $PQ$ be a pair of distinct points, and $d$ any positive constant. There exists in the neighborhood of the point $Q$ at least one point which is joined to $P$ by no two extremals whose $J$-lengths are equal and less than $d$.*

According to Lemma 16.1 there exists in the neighborhood of $Q$ at least one point $Q'$ which is joined to $P$ by no extremals on which $J \leqq d$ and on which $Q'$ is conjugate to $P$. Suppose, however, that there are at least two extremals $g$ and $g'$ joining $P$ to $Q'$ with equal $J$-lengths. Suppose their direction cosines at $Q'$ are $(p)$ and $(q)$.

Since $Q'$ is not conjugate to $P$, a slight variation of $Q'$ will cause a slight variation of the extremals from $P$ to $Q'$ neighboring the initial $g$ and $g'$. The $J$-lengths of these extremals will be analytic functions of the coördinates of $Q'$.

The difference of the $J$-lengths of the extremals $g$ and $g'$ will have partial derivatives given by (4.1), and as seen in §4, not all of these partial derivatives are zero for $(p) \neq (q)$. The locus of points $Q'$ neighboring the initial position of $Q'$ for which two or more of the extremals have $J$-lengths which are equal and less than $d$ or near $d$, will thus lie on a finite number of analytic $(m-1)$-dimensional manifolds without singularities. The lemma follows at once.

With the aid of Lemmas 16.1 and 16.2 one proves the following lemma. The proof is similar to that of Theorem 10.

LEMMA 16.3. *For a fixed point $P$ there exists in the neighborhood of every point $Q \neq P$ at least one point which is not a conjugate point of $P$ nor which is joined to $P$ by more than one extremal of any one $J$-length.*

17. **Example. Geodesics on a knob.** Suppose we have an analytic surface $S$, without singularities, with boundary $B$, and homeomorphic to a circular disc. On $S$ consider the integral of arc length. Suppose $S$ is extremal-convex. Suppose $S$ possesses a knob-shaped protuberance. More exactly suppose there is a portion of $S$ homeomorphic to a circular disc, and bounded by a closed geodesic $g$ that is shorter than nearby closed curves.

Let $R$ be the region between $g$ and $B$. Let $P$ and $Q$ be any pair of non-specialized points on $R$. The geodesics $\gamma$ which give an absolute minimum to the arc length relative to all other admissible curves on $R$ which join $P$ to $Q$, and are deformable into $\gamma$ on $R$, are infinite in number and of type zero. From Theorem 9 we have the following.

*If $P$ and $Q$ are any two non-specialized points on $R$ there will be an infinite set of geodesics of type one joining $P$ to $Q$ on $S$.*

By the use of Theorems 7 and 11 it is not difficult to prove the following more general theorem.

*If $P$ and $Q$ are any two points whatsoever on $R$, there will be an infinite set of geodesics joining $P$ to $Q$, upon each of which there will be at least one point conjugate to $P$.*

18. **The invariance of the connectivities.** We shall first concern ourselves with the dependence of the connectivities of a domain $(a, \rho)$ upon the choice of $n$, the number of vertices in $(\pi)$. It will be convenient to indicate the apparent dependence of $(a, \rho)$ on $n$ by now representing this domain by $(a, \rho, n)$. We shall prove the following.

THEOREM 12. *The connectivity numbers are independent of $n$, in the sense that the connectivity numbers of $(a, \rho, n)$ equal those of $(a, \rho, n')$, where $n$ and $n'$ are any two admissible choices of $n$.*

Let there be given a point $(\pi)'$ on $(a, \rho, n')$, determining a broken extremal $g'$. Let $a(\pi')$ stand for the point $(\pi)$ whose $n$ vertices divide $g'$ into $n+1$ successive segments $h_i$ of equal $J$-length. Let $g$ be the admissible broken extremal determined by $(\pi)$. We shall prove the following statement.

(a) *The curve $g'$ can be continuously deformed on $S$ into the curve $g$ through the mediation of ordinary curves which join $P$ to $Q$, and whose $J$-lengths do not exceed that of $g'$.*

We shall take the time $t$ as the parameter of our deformation, and let it vary from 0 to 1.

Suppose the end points of the segment $h_i$ of $g'$ are joined by an elementary extremal $k_i$ of $g$. For each value of $t$ we suppose $h_i$ divided into two successive segments the ratio of whose $J$-lengths is that of $t$ to $1-t$. For each value of $t$ from 0 to 1 we now replace the second of these segments of $h_i$ by itself, while we replace the first by an elementary extremal that joins its end points. In this manner $h_i$ will be deformed into $k_i$, and thereby $g'$ into $g$. Thus the statement (a) is proved.

Each point $(\pi)$ determines a point $b(\pi) = (\pi)'$ exactly as $(\pi)$ determines $a(\pi') = (\pi)$. Statement (b) will now be proved.

(b) *For every point $(\pi)$ on the domain $(a, \rho, n)$ the point $(\pi)'' = a(b(\pi))$ lies on the domain $(a, \rho, n)$ also. Moreover, there exists a deformation $F$ on $(a, \rho, n)$, of the points $(\pi)$ on $(a, \rho, n)$, which carries these points into the corresponding points $(\pi)''$.*

Let $g$, $g'$ and $g''$ be respectively the broken extremals determined by $(\pi)$, $b(\pi)$ and $(\pi)''$. By deformations similar to those described under (a),

$g$ can be deformed into $g'$, and thence into $g''$ without increasing $J$. Let $t$ be a parameter of the resultant deformation, and vary from 0 to 1. Let $Z(t)$ be the curve that thereby replaces $g$ at the time $t$. A point $U$ on $Z(t)$ may be *determined* by giving $t$ and the $J$-coördinate $u$ of the point $U$ on $Z(t)$. See §5.

A deformation $F$ of the point $(\pi)$ into the corresponding point $(\pi)''$ will now be defined. In the space of the points $(x)$ let each vertex of $(\pi)$ move to the corresponding vertex of $(\pi)''$ in such a manner that the pair $(t, u)$ which *determines* (see above) this variable vertex moves on a straight line in the $(t, u)$ plane at a constant velocity equal to the distance to be traversed. The corresponding deformation $F$ is readily seen to have the desired properties.

If $C_j$ is any complex on the domain $(a, \rho, n)$ we shall denote by $b(C_j)$ that complex on $(a, \rho, n')$ which consists of the images $b(\pi)$ of points $(\pi)$ on $C_j$. Reciprocally if $C_j'$ is a complex on $(a, \rho, n')$, $a(C_j')$ will denote the set of images $a(\pi')$ of points $(\pi)'$ on $C_j'$. With this understood we now prove a final statement (c).

(c) *If the cycles*

$$C_j^1, \cdots, C_j^r \tag{18.1}$$

*form a complete $j$-set for $(a, \rho, n')$, the cycles*

$$a(C_j^1), \cdots, a(C_j^r) \tag{18.2}$$

*will form a complete $j$-set for the domain $(a, \rho, n)$.*

Let $C_j$ be any $j$-cycle on the domain $(a, \rho, n)$. Then $b(C_j)$ will be a $j$-cycle on the domain $(a, \rho, n')$. Because (18.1) gives a complete $j$-set for $(a, \rho, n')$ we have

$$b(C_j) + \Sigma C_j^i \equiv C_{j+1} \quad \text{on } (a, \rho, n'), \tag{18.3}$$

where $\Sigma$ stands for a suitable sum of cycles (18.1) and $C_{j+1}$ is a complex on $(a, \rho, n')$. From (18.3) we see that

$$a(b(C_j)) + \Sigma a(C_j^i) \equiv a(C_{j+1}) \quad \text{on } (a, \rho, n). \tag{18.4}$$

From (b) we see however that

$$a(b(C_j)) \sim C_j \quad \text{on } (a, \rho, n). \tag{18.5}$$

From (18.4) and (18.5) we see finally that

$$C_j \sim \Sigma a(C_j^i) \quad \text{on } (a, \rho, n). \tag{18.6}$$

Thus the $j$th connectivity number of $(a, \rho, n)$ will be at most $r$. We can reverse the rôles of $(a, \rho, n)$ and $(a, \rho, n')$. We infer then that the con-

nectivities of $(a, \rho, n)$ and $(a, \rho, n')$ are the same. Thus (c) and the theorem are proved.

We now give another theorem on the invariance of the connectivities.

THEOREM 13. *The connectivities of the domain $(a, \rho)$ remain unchanged during any continuous variation of the end points $P$ and $Q$ on $S$ $(P \neq Q)$, provided the constant $a$ remain a non-critical value of $J$.*

Let the domain $(a, \rho)$ of admissible points $(\pi)$, set up for a pair of points, $PQ$, now be indicated by

$$(a, \rho, P, Q). \tag{18.7}$$

If $e$ be a sufficiently small positive constant the domain (18.7) can be $J$-deformed onto the domain (see § 6)

$$(a - e, \rho - e, P, Q). \tag{18.8}$$

If $e_1$ now be chosen as a positive constant less than $e$, and $P'Q'$ be a pair of points sufficiently near $PQ$, the domain

$$(a - e_1, \rho - e, P', Q') \tag{18.9}$$

will be included in the domain (18.7) and will include the domain (18.8).

Now the domain (18.7) can be $J$-deformed on itself into a set of points on (18.8), and hence into a set of points on (18.9). It follows that the connectivities of (18.7) equal those of (18.9).

This remains true for any smaller choice of $e_1$ and choice of $P'Q'$ sufficiently near $PQ$. But if $e_1$ be sufficiently small, and $P'Q'$ so near $PQ$ that $a$ remains a non-critical value of $J$, the domain

$$(a, \rho, P', Q') \tag{18.10}$$

can be $J$-deformed on itself into a set of points on (18.9) and hence has the connectivities of (18.9). Hence (18.10) and (18.7) have the same connectivities, and the theorem is proved.

**19. Extremals on closed, regular, analytic manifolds.** All of the previous developments go through for this case as in the case of the region $S$, except for obvious changes of which we will enumerate the most important.

All references to the boundary, or the extremal-convex hypothesis are to be omitted.

Instead of having one set of coördinates $(x)$, it will in general be necessary to give the manifolds by a finite set of overlapping parametric representations.

All the extremals through a point $P$ may go through a second point $Q$.

This case should be treated separately. It can never occur for a non-specialized pair of points.

The theory of extremals on elementary regions will apply to elementary regions on the manifold. The relations between the extremals and the connectivities of the domain $(a, \rho)$ are given by Theorem 6.

The most important case is the case where the manifold is homeomorphic with an $m$-sphere. The following theorem will be proved in a later paper.

THEOREM 14. *On a manifold homeomorphic with an m-sphere there are infinitely many extremals joining any two fixed points, including extremals of arbitrarily great length, with arbitrarily many conjugate points on them.*

In this theorem it is understood that two extremals are counted as different if they have different lengths, even if they overlap. The proof of this theorem depends upon the preceding work together with a method of determining the connectivities $R_i$, called *the topological continuation of extremals.*

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# A CONVERSE OF THE JORDAN-BROUWER SEPARATION THEOREM IN THREE DIMENSIONS*

BY

R. L. WILDER

That the closed $(n-1)$-manifold immersed in euclidean space, $E_n$, of $n$ dimensions $(n>2)$, separates $E_n$ into just two domains of which it is the common boundary, was shown by Brouwer in 1912.† That the points of the manifold are accessible from each of its complementary domains Brouwer proved in an accompanying paper,‡ and in the latter connection he gave an example to show that a bounded and closed point set which separates $E_n$ into just two domains and every point of which is accessible from each of these domains, is not necessarily homeomorphic with a closed manifold.

The above results of Brouwer, in so far as the connectivity of the set residual to the manifold in $E_n$ is concerned, received considerable extension at the hands of J. W. Alexander, who not only demonstrated, using modulo 2 Betti numbers, that the residual set is just two connected domains, but that a certain duality exists between the connectivity numbers of the set and those of its complement.§ If we denote the $i$th connectivity number of a set $F$ by $R^i(F)$, then, for the particular case where $C^i$ is a set in $E_n$ homeomorphic with an $i$-sphere, Alexander showed that

$$R^i(C^i) = R^{n-i-1}(E_n - C^i) = 2, \tag{1}$$

$$R^s(C^i) = R^{n-s-1}(E_n - C^i) = 1 \qquad (s \neq i). \tag{2}$$

In the case where $n=3$ and $i=2$, relation (1) states that the complement of a simple closed surface in $E_3$ is just two domains (as also shown by Brouwer),

* Presented to the Society, December 31, 1928, March 29, 1929, and August 30, 1929; received by the editors October 11, 1929.

† L. E. J. Brouwer, *Beweis des Jordanschen Satzes für den n-dimensionalen Raum*, Mathematische Annalen, vol. 71 (1912), pp. 314–327.

‡ L. E. J. Brouwer, *Über Jordansche Mannigfaltigkeiten*, ibid., pp. 320–327.

§ J. W. Alexander, *A proof and extension of the Jordan-Brouwer separation theorem*, these Transactions, vol. 23 (1922), pp. 333–349. Familiarity with this work is assumed throughout the present paper.

The terminology "Betti number" is at present being employed more and more to denote connectivity numbers that are uniformly less by one unit than the numbers $R^i$ of the present paper. However, since the present paper is based so much upon the above paper of Alexander, it seems advisable to retain the connectivity numbers as defined in that connection (although they are not termed "Betti numbers" by Alexander).

and relation (2) states the important additional fact that every closed $s$-chain (where $s<2$) in the complement bounds an $(s+1)$-chain in the complement. Thus, although as Alexander pointed out in another connection* the domains complementary to a simple closed surface in $E_3$ may not be simply connected (may, in fact, require an infinite number of generators), every closed 1-chain in either of the domains bounds in that domain.

In contemplating so simple a surface as the torus, it becomes apparent that in formulating a *converse* of the Jordan-Brouwer separation theorem the relation (2) must be taken into account as well as relation (1). On the other hand, a glance at Brouwer's example cited above is sufficient to inform one that even with the duality relations (1) and (2), the accessibility of the points of a surface from each of its complementary domains is not enough to ensure that surface being, in $E_3$, a simple closed surface. Some condition must be imposed that will do away with the crinkliness of the surface. It occurred to the author that such a condition might be the following: If $K$ is a simple closed surface in $E_3$, $D$ one of the domains complementary to $K$, and $\epsilon$ a positive number, then there exists a positive number $\delta$ such that if $\Gamma^0$ is a closed 0-chain in $D$ of diameter less than $\delta$, then $\Gamma^0$ bounds an open 1-chain $\Gamma^1$ in $D$ whose diameter is less than $\epsilon$. This property of $D$ may be concisely expressed by saying that *the 0-chains of $D$ are uniformly homologous to zero in $D$.*

It will be shown in §1 of this paper (which may be considered as complementary to Alexander's paper on the Jordan-Brouwer separation theorem, since it utilizes the machinery introduced by him in order to treat accessibility properties into which he did not go), that the above condition is a necessary condition for the general $n$-dimensional case, and in §3 it will be shown that together with part of Alexander's duality relation it enables one to give, in $E_3$, a converse of the Jordan-Brouwer theorem.

The chief difficulty encountered by the author in applying the above, essentially combinatorial, properties,† to characterizing the sphere immersed in $E_3$, was that the known characterizations of the simple closed surface‡

* J. W. Alexander, *An example of a simply connected surface bounding a region which is not simply connected*, Proceedings of the National Academy of Sciences, vol. 10 (1924), pp. 8–10.

† It will be shown below, in the Appendix, that *uniformly homologous to zero* as a condition relative to the 0-chains of a domain is equivalent to the condition *uniform connectedness im kleinen* for the domain.

‡ The earliest characterization of which the author knows was announced by R. L. Moore and J. R. Kline, their abstract being in the Bulletin of the American Mathematical Society, vol. 28 (1922), p. 380. The same definition is given by Miss I. Gawehn (except for the greater generality of the space considered), in *Über unberandete 2-dimensionale Mannigfaltigkeiten*, Mathematische Annalen, vol. 98 (1927–28), pp. 321–354. Cf. also C. Kuratowski, *Une caractérisation topologique de la surface de la sphère*, Fundamenta Mathematicae, vol. 13 (1929), pp. 307–318, as well as abstracts by L. Zippin, Bulletin of the American Mathematical Society, vol. 35 (1929), p. 154 and p. 293.

were all of a topological type which did not seem to lend themselves very readily to association with the combinatorial properties of the complement. Accordingly a new characterization of the simple closed surface by means of internal properties that are readily associated with the combinatorial properties of the complement was worked out, and is given below in §2.

In closing this introduction the author wishes to call to the attention of the reader the problem of giving a converse of the Jordan-Brouwer separation theorem in $E_3$ by certain accessibility conditions that are more in the spirit of the Schoenflies converse as formulated in $E_2$,* and it is hoped that if the attempt is made to find such conditions, the present work will be of some assistance as having already broken down some of the barriers. For in the opinion of the author the analysis situs relations between closed sets and their complements in spaces of three and higher dimensions will be most easily discovered by first having regard for the connectivity numbers of the complements,† and although it would be unwise to make any predictions, it would certainly seem as though to disregard these in the case of the $(n-1)$-sphere, when already known, in the attempt to obtain equivalent accessibility conditions, would involve an extreme waste of energy.

1. We shall show in this section‡ that if $M^{n-1}$ is an $(n-1)$-sphere im-

* In this connection see J. R. Kline, *Separation theorems and their relation to recent developments in analysis situs*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 155–192. See especially pp. 156–159 and top of p. 191. As will be observed later on, *regular* accessibility is an immediate consequence of the conditions given for the converse in the present paper.

† The recent work of P. Alexandroff bears out this opinion strikingly. Cf. his *Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension*, Annals of Mathematics, (2), vol. 30 (1928), pp. 101–187. The author wishes to seize this opportunity to express his indebtedness to Dr. Alexandroff, contact with whom, both through his memoirs and personally (during a recent brief visit at the University of Michigan), directed the attention of the author to the possibility of associating continuity and combinatorial methods in treating the problems of analysis situs.

‡ The content of this section was presented, along with certain other results, to this Society, March 29, 1929, under separate title (cf. Bulletin of the American Mathematical Society, vol. 35, p. 458, abstract No. 22). It is felt, however, that the theorems and corollaries given above properly belong in the present paper. As for the other results of the paper just referred to, the theorem that the points of an $i$-cell $(0 \leqq i \leqq n-1)$ are regularly accessible from the complement in $E_n$ is a special case of a more recent theorem established by the author, to the effect that in $E_n$ the points of any closed set which is homeomorphic with a subset of $E_{n-1}$ are regularly accessible from the complement; several of the other results were found to duplicate already known theorems and were withdrawn from the announcement in abstract. One of the things so deleted was an example which the author believes to be not without interest, and which he has not seen in the literature, viz., an arc in $E_3$ which cannot lie on, or bound, any set homeomorphic with a 2-cell, and which cannot be thrown into a straight line interval by a (1-1)-continuous transformation of $E_3$ into itself. (Cf. L. Antoine, *Sur l'homéomorphie de deux figures et de leurs voisinages*, Journal de Mathématiques, (8), vol. 4 (1921), pp. 221–325.) The chief interest of the example is its ease of description: On a great circle of the 2-sphere, let $ABCDEA$ be points in the cyclical order named, and let the portion $BC$ of the

mersed in $E_n$,* and $D$ is either of its complementary domains, then the 0-chains of $D$ are uniformly homologous to zero in $D$. Consequently the domain $D$ is uniformly connected im kleinen (see Appendix)—a fact which was established for the case where $n=2$ and $M^{n-1}$ is a 1-sphere by R. L. Moore.† As a corollary it follows that the points of the sphere are not only accessible, but are regularly accessible, from the complementary domains.

THEOREM 1. *Let $M^{n-1}$ be an $(n-1)$-sphere immersed in $E_n$, and let $D$ be one of the domains of $E_n - M^{n-1}$. Then the 0-chains of $D$ are uniformly homologous to zero in $D$.*

Let the domain $E_n-(M^{n-1}+D)$ be denoted by $D_1$.

Suppose the 0-chains in $D$ are not uniformly homologous to zero. Then it easily follows that there is a point $P$ on $M^{n-1}$ and a spherical neighborhood $U$ of $P$, such that in every neighborhood of $P$ there is a 0-chain which does not bound in that part of $D$ which lies in $U$. We shall denote the frontier of $U$ by $F$.

There is a cell $C^{n-1}$ of $M^{n-1}$ which contains $P$ as a non-boundary point and lies wholly in $U$. Denote the boundary of $C^{n-1}$ by $C^{n-2}$, and let the set of all those points of $M^{n-1}$ that are not interior points of $C^{n-1}$ be denoted by $B^{n-1}$. Let $V$ be a spherical neighborhood of $P$ of such a radius that $V$ contains no points of $B^{n-1}$. Let $x_1^0$ and $x_2^0$ be any two 0-cells of $D$ that lie in $V$. Since the closed chain $x_1^0+x_2^0$ bounds a 1-chain $L_B^1$ interior to $V$, we have

$$x_1^0 + x_2^0 \sim 0 \qquad [\text{mod } 2,\ E_n - (F + B^{n-1})]. \tag{3}$$

In order to deal with a specific case, we may assume that a 0-cell, $y_1^0$, of $L_B^1$, lies in $D_1$, $L_B^1$ having been obtained, say, by joining $x_1^0$ and $x_2^0$ to such a cell by chains $L_{B_1}^1$ and $L_{B_2}^1$, respectively.

---

arc $ABCD$ be replaced by a new portion containing two infinite sequences of simple trefoil knots (i.e., knots in the sense that the great circle becomes knotted), which of course have to leave the 2-sphere, and which respectively approach the points $B$ and $C$ as limiting points. This can be done in such a way that the new portion $BC$ is still an arc from $B$ to $C$ that does not meet the arc $BAED$ of the great circle. Then the portions $AB$ and $CD$ of the great circle, together with the new portion $BC$, is an arc of the type described above.

* I.e., $M^{n-1}$ is a point set which is in (1-1)-continuous correspondence with an $(n-1)$-sphere, and is therefore not necessarily formed by cells of the subdivisions of $E_n$ which are used in the congruences and homologies of the proof. Cf. J. W. Alexander, these Transactions, loc. cit., for definitions of *chain*, *homology*, etc. The reader should be on his guard in the matter of terminology which has various meanings depending upon its application in a point-set theoretic sense or in a combinatorial sense. (E.g., "chain," "closed," "open," etc. Present usage seems to apply the term "chain" only to open chains, chains that are closed being termed "cycles." We are retaining the terminology of Alexander's paper, however, in the present connection.)

† R. L. Moore, *A characterization of Jordan regions by properties having no reference to their boundaries*, Proceedings of the National Academy of Sciences, vol. 4 (1918), pp. 364–370.

On the other hand, as every 0-chain in $D$ bounds in $D$, even though it may not bound im kleinen, there is a chain $L_C{}^1$ in $D$ which is bounded by the chain $x_1{}^0+x_2{}^0$. Hence,

$$x_1{}^0 + x_2{}^0 \sim 0 \qquad (\text{mod } 2,\ E_n - C^{n-1}). \tag{4}$$

If now we can show that the closed chain

$$L_B{}^1 + L_C{}^1 \sim 0 \qquad (\text{mod } 2,\ E_n - C^{n-2}), \tag{5}$$

it will follow from relations (3), (4) and (5), and Alexander's Corollary $W^i$,* that $x_1{}^0+x_2{}^0$ bounds in the set common to $D$ and $U$, and since no restriction is placed upon the choice of $x_1{}^0$ and $x_2{}^0$, a contradiction of the supposition that the 0-chains of $D$ are not uniformly homologous to zero will be obtained. We therefore proceed to demonstrate the validity of relation (5).

Let $C_B{}^{n-1}$ be an arbitrary cell of $M^{n-1}$ interior to $B^{n-1}$. Then $x_1{}^0$ and $y_1{}^0$ bound a 1-chain $K_1{}^1$ in $E_n-(M^{n-1}-C_B{}^{n-1})$. The closed chain $K_1{}^1+L_{B1}{}^1$ links $C^{n-2}$, for if it did not, then, by virtue of Alexander's Corollary $W^i$, $x_1{}^0+y_1{}^0$ would bound in $E_n-M^{n-1}$. Similarly, the closed chain $K_1{}^1+L_C{}^1+L_{B2}{}^1$ links $C^{n-2}$. However, by relation (1) there is only one linearly independent non-bounding 1-chain in $E_n-C^{n-2}$. Hence

$$K_1{}^1 + L_{B1}{}^1 \sim K_1{}^1 + L_C{}^1 + L_{B2}{}^1 \qquad (\text{mod } 2,\ E - C^{n-2}). \tag{6}$$

Since homology (6) implies homology (5), the theorem is proved.

COROLLARY 1. *The points of an $(n-1)$-sphere, $M^{n-1}$, immersed in $E_n$, are regularly accessible† from the complement of the sphere.*

Let $P$ be a point of $M^{n-1}$, and $D$ one of the domains complementary to $M^{n-1}$. Since there are only two domains in the complement of $M^{n-1}$, it will be sufficient to prove $P$ regularly accessible from $D$.

By virtue of Theorem 1 there exists (1) a sequence of spherical neighborhoods of $P$, $U_1, U_2, U_3, \cdots$, such that if $r_k$ is the radius of $U_k$, then $r_k>r_{k+1}$, and $\lim_{k\to\infty} r_k=0$; and such that for $k>1$, any closed 0-chain of $D$ that lies in $U_k$ bounds a 1-chain that lies wholly in the common part of $U_{k-1}$ and $D$; (2) a sequence of 0-cells $x_2{}^0, x_3{}^0, x_4{}^0, \cdots$, such that, for every $k$, $x_k{}^0$ lies in the common part of $D$ and $U_k$.

* These Transactions, loc. cit., p. 342.

† A point $P$ is said to be *regularly accessible* from a point set $D$ if for every positive number $\epsilon$ there exists a positive number $\delta$ such that if $Q$ is a point of $D$ whose distance from $P$ is less than $\delta$, then there is an arc $PQ$ in $D+P$ whose diameter is less than $\epsilon$. Cf. G. T. Whyburn, *Concerning the open subsets of a plane continuous curve*, Proceedings of the National Academy of Sciences, vol. 13 (1927), pp. 650–657.

For any $k$, $x_k^0$ and $x_{k+1}^0$ bound a 1-chain, $\Gamma_k^1$, which lies wholly in the common part of $D$ and $U_{k-1}$. As the set of points

$$M = \sum_{k=2}^{\infty} \Gamma_k^1 + P$$

is a continuous curve, there is an arc from $x_2^0$ to $P$ that lies in $M$, and hence (except for $P$) in the common part of $U_1$ and $D$. As the first neighborhood, $U_1$, is arbitrary, and as in every neighborhood of a point $x$ of $D$ in $U_2$ there can be found a 0-cell $x_2^0$ in $D$, the corollary is proved.

If, in the proof of Theorem 1, $M^{n-1}$ denotes an $(n-1)$-manifold, and $P$ denotes a point which is interior to (i.e., not on the boundary of) a cell $C^{n-1}$ which is one of the cells defining $M^{n-1}$, then the argument to show that $x_1^0 + x_2^0$ is homologous to zero in $D \cdot U$ goes through as before. However, if $P$ is such a point of the manifold that it cannot lie interior to any such cell, then the above 0-chain may fail to bound in $D \cdot U$. (Cf., for instance, the set $K$ defined in the Appendix, paragraph beginning "Condition (3)".) Hence we can state the following theorem:

THEOREM 1a. *Let $M^{n-1}$ be any $(n-1)$-manifold immersed in $E_n$, $D$ a domain complementary to $M^{n-1}$, and $P$ a point of $M^{n-1}$ which can be considered as interior to an $(n-1)$-cell of definition of $M^{n-1}$. Then, if $C^{n-1}$ is any such cell, and $U$ is any spherical neighborhood of $P$ that encloses only points of $M^{n-1}$ that are interior to $C^{n-1}$, there exists a neighborhood $V$ of $P$ such that any 0-chain in $D \cdot V$ bounds a 1-chain in $D \cdot U$. If every point of $M^{n-1}$ can be considered as lying interior to an $(n-1)$-cell of definition of $M^{n-1}$, the 0-chains of $D$ are uniformly homologous to zero in $D$.**

COROLLARY 2. *The points of an $(n-1)$-manifold, $M^{n-1}$, immersed in $E_n$, are regularly accessible from the complement of the manifold.*

If $P$ is an interior point of a cell $C^{n-1}$ of $M^{n-1}$, the proof is the same as that of the above Corollary 1. If not, then $P$ is on the boundaries of a certain number (at least four) of $(n-1)$-cells of $M^{n-1}$. Let $V$ be any neighborhood of $P$. Since any $(n-1)$-cell of an $(n-1)$-manifold is homeomorphic with an $(n-2)$-sphere and its interior in $E_{n-1}$, there is a neighborhood $U$ of $P$ such that any point of $M^{n-1}$ which lies in $U$ and is interior to a cell $C^{n-1}$ of $M^{n-1}$

** In order to avoid confusion in the mind of the reader, it should perhaps be pointed out here that we are using the definition of manifold as given by Alexander in his paper on the Jordan-Brouwer Separation Theorem. That is, an $i$-manifold is a closed $i$-chain whose $i$th connectivity number has the value 2. Some authors (cf., for instance, Veblen's *Analysis Situs*, published by this Society, 1922, p. 88) state the additional condition that every point of the $i$-manifold shall have a neighborhood, relative to the manifold, which is an $i$-cell. The definition as given by Alexander of course yields greater generality to the results obtained.

is joined to $P$ by an arc of $M^{n-1} \cdot V$ that lies, except for $P$, wholly interior to $C^{n-1}$ (by Corollary 1). Let $Q$ be any point of $D$ (one of the domains of $E_n - M^{n-1}$) that lies in $U$. We shall show that there is an arc from $P$ to $Q$ lying, except for $P$, wholly in $D \cdot V$.

Let $PQ$ be any arc of $U$ from $P$ to $Q$, and let $P_1$ be its first point on $M^{n-1}$, in the order from $Q$ to $P$. Since in any neighborhood of a point of $M^{n-1}$ there are points of $M^{n-1}$ not on the boundaries of its $(n-1)$-cells, we may assume that $P_1$ is interior to a cell $C^{n-1}$ of $M^{n-1}$ whose boundary, $C^{n-2}$, contains $P$. There is an arc, $t$, of $M^{n-1}$, from $P_1$ to $P$ and lying wholly interior to $C^{n-1}$, except for $P$, as well as interior to $V$. Let $P_2, P_3, P_4, \cdots$ be a sequence of points of $t$ having $P$ as a sequential limit point, and let $\epsilon$ denote the distance between $t$ and the frontier of $V$ as ordinarily defined for point sets.

Consider the portion $P_1P_2$ of $t$. Since all points of $P_1P_2$ are interior points of $C^{n-1}$, it is easy to show that there lies, in $D \cdot V$, an arc $Q_1Q_2$ approximating $P_1P_2$, where $Q_1$ is a point of $QP_1$ in the neighborhood of $P_1$, and $Q_2$ a point in the neighborhood of $P_2$. For if $x$ is any point of $P_1P_2$, there is a number $\delta$ such that any 0-chain of $D$ in $S(x, \delta)$ bounds a 1-chain of $D$ in $S(x, \frac{1}{2}\epsilon)$ (Theorem 1a). A simple chain* of the neighborhoods $S(x, \delta)$ from $P_1$ to $P_2$, together with the joining of 0-cells in the successive links of the simple chain, yields the arc $Q_1Q_2$ desired.

By successive approximations to the arcs $P_iP_{i+1}$ by arcs $Q_iQ_{i+1}$ in $D \cdot V$ in such a manner that if $x_i$ is any point of $Q_iQ_{i+1}$, $\lim_{i=\infty} x_i = P$, it is clear that the set $QQ_1 + \sum_{i=1}^{\infty} Q_iQ_{i+1} + P$ is a continuous curve containing an arc from $Q$ to $P$ and lying, except for $P$, wholly in $D \cdot V$.

2.† We shall now prove the following theorem:

THEOREM 2. *In a metric separable space let $M$ be a compact set containing at least one simple closed curve, and satisfying the following conditions*: (1) *if $t$ is an arc of $M$, then $M - t$ is connected*; (2) *if $J$ is a simple closed curve of $M$, then $M - J$ is the sum of two uniformly connected im kleinen‡ components. Then $M$ is a simple closed surface*; *i.e., a set homeomorphic with the sphere* $x^2 + y^2 + z^2 = 1$.§

* Cf. R. L. Moore, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), pp. 131–164. In particular, see bottom of p. 134, and Theorem 10 of this paper.

† The content of this section was presented to the Society December 31, 1928.

‡ A set of points $K$ is said to be *uniformly connected im kleinen* if for every $\epsilon > 0$ there exists a $\delta > 0$, such that if $P$ and $Q$ are points of $K$ whose distance apart is less than $\delta$, then $P$ and $Q$ are in a connected subset of $K$ whose diameter is less than $\epsilon$. However, for an open subset of a continuous curve, it is easy to prove that the words "a connected subset" in this definition may be replaced by the words "an arc."

§ When this paper was in process of completion, Dr. Leo Zippin communicated to the author that he had succeeded in showing that a continuous curve $C$ which has the properties that (1) if

For ease of reference, we shall divide the proof into sections:

I. *The set $M$ is connected im kleinen.* For if $J$ is a simple closed curve of $M$, and $M_1$ and $M_2$ are the components of $M-J$, then $M_1+J$ and $M_2+J$ are connected im kleinen by a theorem of R. L. Moore.* Accordingly the set $M$ is itself connected im kleinen.

II. *If $J$ is a simple closed curve of $M$, and $M_1$ and $M_2$ are the components of $M-J$, then $J$ is identical with the boundary of each of the sets $M_1$, $M_2$.* For suppose $J$ contains a point, $P$, which is not a limit point of $M_1$, say. Then an arc, $t$, of $J$, contains all the boundary points (if any) of $M_1$. Then

$$M - t = M_1 + (M_2 + J - t).$$

Since $M_1$ and $M_2+J-t$ are mutually separated sets, a violation of condition (1) results.

III. *The set $M$ is connected.* For $M$ contains at least one simple closed curve, $J$, and, by II, $J$ is the common boundary of the components of $M-J$.

IV. *If $P$ is a point of $M$, then there is an arc of $M$ which contains $P$, and of which $P$ is not an end point.*†

Since $M$ is a continuous curve, being a connected im kleinen, closed and connected set, there is an arc $PP'$ of $M$, joining $P$ to some point $P'$ of $M-P$. By condition (1) the set $M-PP'=D$ is connected. If $P$ is not a limit point of $D$, then there is a point $Q$ of $PP'$ distinct from both $P$ and $P'$, such that the portion $PQ$ of $PP'$ contains no limit points of $D$. Then the set

$$M - P'Q = D + (PQ - Q)$$

is not connected, since $D$ and $PQ-Q$ are mutually separated sets. But this constitutes a violation of condition (1). Consequently $D$ has $P$ as a limit point.

The set $D+P$ is connected im kleinen. If not, then $P$ is the only point at which it is not connected im kleinen, since $D$ is an open subset of $M$. Then there exists, by Theorem 1 of my paper *Characterizations of continuous curves that are perfectly continuous*,‡ a positive number $\epsilon$ such that for every positive

---

$t$ is an arc of $C$, then $C-t$ is connected, (2) if $J$ is a simple closed curve of $C$, then $C-J$ is not connected, is a simple closed surface. This definition is an obvious improvement over that given above, and by its use the last two parts of the proof of the theorem of §3 can be omitted.

* R. L. Moore, *Concerning connectedness im kleinen and a related property*, Fundamenta Mathematicae, vol. 3 (1922), pp. 232–237, Theorem 1. The proof given by Moore evidently holds for a compact set in any topological space.

† Note added in proof: Since this paper was submitted to these Transactions, a paper by W. L. Ayres has appeared in the American Journal of Mathematics for October, 1929, in which it is shown (Theorem 10) that if a continuous curve in $E_n$ contains no cut point, then the curve is cyclicly connected. As our curve $M$ clearly contains no cut point by virtue of condition (1), it is obvious that the results IV and V follow from Ayres' theorem.

‡ Proceedings of the National Academy of Sciences, vol. 15 (1929), pp. 614–621.

number $\delta < \epsilon$ there is a point $Q$ of $D$ whose distance from $P$ is less than $\delta$ and which is not in the same quasi-component with $P$ of the set of points $(D+P)\cdot S(P, \epsilon)$.* It follows at once that there is a sequence of points $x_1, x_2, x_3, \cdots$, such that $\epsilon/2 > \rho(x_n, P) > \epsilon/3$, and such that no two points of this sequence lie in the same quasi-component of $(D+P)\cdot S(P, \epsilon)$. As $M$ is compact, there is a limit point, $x$, of the set of points $\sum_{n=1}^{\infty} x_n$. The point $x$ cannot lie in $D$, since $D$ is connected im kleinen. Hence $x$ is on the arc $PP'$, and since there is no loss of generality in assuming that $\rho(P, P') > \epsilon$, we can say that $x$ is distinct from both $P$ and $P'$.

However, $D$ is uniformly connected im kleinen in the neighborhood of $x$. For let $t$ be an arc in $PP'$ whose end points are distinct from $P$ and $P'$, and which contains $x$. Since $M-t$ is connected, by condition (1), there exists in $M-t$, by a theorem of R. L. Moore,† an arc from $P$ to $P'$, and this arc together with the arc $PP'$ contains a simple closed curve $J$ which contains $t$, consisting of two arcs, $AxB$ of $PP'$ and $AyB$ of $D+A+B$. Let the components of $M-J$ be $M_1$ and $M_2$, and let $\eta$ be a positive number such that $S(x, \eta)$ contains no points of $J+PP'$ that are not points of $t$. Then by condition (2) there is a positive number $\rho$ such that every two points of $M_1$ that lie in $S(x, \rho)$ are joined by an arc of $M_1$ in $S(x, \eta)$, and a similar statement holds for $M_2$. Since all points of $M-J$ in $S(P, \eta)$ belong to $D$, it is clear that if $\eta$ is taken so small that $S(x, \eta)$ is a subset of $S(P, \epsilon)$, the points $x_1, x_2, x_3, \cdots$ cannot all lie in distinct quasi-components of $(D+P)\cdot S(P, \epsilon)$.

Accordingly $D+P$ must be connected im kleinen, and by a theorem of Whyburn‡ $P$ is regularly accessible from $D$. Hence there is an arc of $M$ containing $P$ and of which $P$ is not an end point.

V. *If $P$ is a point of $M$, then $P$ is on some simple closed curve of $M$.* This follows at once from IV, and the method used in IV to obtain the simple closed curve $J$.

VI. *If $P$ is a point of $M$, $M-P$ is connected.* By V, $P$ is on a simple closed curve $J$ of $M$, and by condition (2) $M-J$ is the sum of two components which have, by II, $J$ as common boundary. Since $M-P$ is the sum of these

---

* Hereafter, if $P$ is a point and $\epsilon$ any positive number, the symbol $S(P, \epsilon)$ will denote a spherical neighborhood of $P$ of radius $\epsilon$; i.e., the set of all points of the space whose distance from $P$ is less than $\epsilon$. In accordance with the usual notation, $M\cdot N$ denotes the set of points common to $M$ and $N$. By $\rho(P, Q)$ we denote the distance between two points $P$ and $Q$.

† R. L. Moore, *Concerning continuous curves in the plane*, Mathematische Zeitschrift, vol. 15 (1922), pp. 254–260, Theorem 1. The theorem is evidently true for the case of a continuous curve in the general space we are considering.

‡ G. T. Whyburn, *Concerning accessibility in the plane and regular accessibility in n dimensions*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 504–510, Theorem 3. Although stated for $E_n$, the proof given for this theorem clearly holds in any metric space.

components together with some of their common boundary points, it is clear that $M-P$ is connected.

VII. *If $M_1$, $M_2$ and $J$ are defined as in* II, *and $AxB$ is an arc such that $<AxB>$* is in $M_1$ and $A+B$ on $J$, then $M_1-<AxB>$ is the sum of two components $R_1$ and $R_2$ whose boundaries are, respectively, $AxB$+an arc $AyB$ of $J$, and $AxB$+an arc $AzB$ of $J$, where $AyB$ and $AzB$ have only $A$ and $B$ in common.*

Let the two arcs into which $A$ and $B$ divide $J$ be $AyB$ and $AzB$. Let $J_1$ and $J_2$ be simple closed curves defined as follows:

$$J_1 = AxB + AyB,$$
$$J_2 = AxB + AzB.$$

By condition (2), $M-J_1$ is the sum of two components $R_1$ and $R_1'$ whose common boundary, by II, is $J_1$. Let $R_1$ be that one of these components which lies in $M_1$. (It is clear that $M_2$ lies in one of the components of $M-J_1$, say in $R_1'$, and that $R_1'$ must therefore contain $<AzB>$, thus requiring that $R_1$ be a subset of $M_1$.) Let that one of the components of $M-J_2$ that lies in $M_1$ be denoted by $R_2$.

Suppose that $M_1$ contains a point $P$ that is not in the set $R_1+R_2+<AxB>$. Then $P$ lies in $R_1'$, and if $P_1$ is a point of $M_2$, there is an arc $PP_1$, from $P$ to $P_1$, in $R_1'$. If $Q$ is the first point of $J$ on this arc, in the order from $P$ to $P_1$, then $PQ-Q$ lies in $M_1$. It is clear that $Q$ is in $<AzB>$; let $S(Q, \epsilon)$ be a neighborhood of $Q$ that does not enclose any point of $AxB$. There is a neighborhood $S(Q, \delta)$ such that any two points of $M_1 \cdot S(Q, \delta)$ are joined in $S(Q, \epsilon)$ by an arc of $M_1$. Since points of $PQ-Q$ and $R_2$ both lie in $S(Q, \delta)$, and such points are then joined by connected subsets of $M-J_2$, a contradiction results from the supposition that $M_1$ contains points not in the set $R_1+R_2+<AxB>$. Hence $M_1=R_1+R_2+<AxB>$.

VIII. *If $P$ is a point of $M$, and $\epsilon$ is any positive number, there is a simple closed curve $J$ in $M$ which does not contain $P$ and such that both $J$ and the component of $M-J$ which contains $P$ lie in $S(P, \epsilon)$.*

By V, $P$ is on some simple closed curve $J$ of $M$. Denote the components of $M-J$ by $M_1$ and $M_2$. We can, of course, suppose that $J$ is not wholly in $S(P, \epsilon)$, and that both $M_1$ and $M_2$ have points in the exterior of $S(P, \epsilon)$.

By virtue of the uniform connectedness im kleinen of $M_1$ and $M_2$, together with II, it is easy to show the existence, in $S(P, \epsilon)$, of arcs $A_1xB_1$ and $A_2yB_2$, where $<A_1xB_1>$ and $<A_2yB_2>$ lie in $M_1$ and $M_2$, respectively, and the points $A_1$, $A_2$, $B_1$, $B_2$ lie on $J$ in the order $A_1A_2PB_2B_1A_1$, and such that the arc $A_1PB_1$ of $J$ lies in $S(P, \epsilon)$. Let the simple closed curve $A_1xB_1+B_1B_2+A_2yB_2$

* By $<AxB>$ we denote the set $AxB-A-B$.

$+A_1A_2$ (where $A_1A_2$ and $B_1B_2$ are portions of $A_1PB_1$ on $J$) be denoted by $J_1$. Let that component of $M-J_1$ that contains $P$ be denoted by $R_1$.

By VII, $M_1-<A_1xB_1>$ is the sum of two components, one of which is bounded by the simple closed curve $T_1=A_1PB_1+A_1xB_1$; denote this component by $K_1$. Similarly, that component of $M_2$ bounded by the simple closed curve $T_2=A_2PB_2+A_2yB_2$ denote by $K_2$. Then

$$R_1 = K_1 + K_2 + < A_2PB_2 > .$$

For by VII, $R_1-<A_2PB_2>$ is the sum of two components $H_1$ and $H_2$, bounded, respectively, by $T_1$ and $T_2$. Now $T_1$ bounds only two components, one of which contains $M_2$. If $H_1$ were the latter component, then $H_1$ would contain $<A_2yB_2>$, which is clearly impossible. Then $H_1$ is that component of $M-T_1$ that does not contain $M_2$, i.e., $K_1$. Similarly, $H_2=K_2$. We note that if $\alpha$ is a point of $M_1$ that is not in $K_1+T_1$, then $R_1$ does not contain $\alpha$.

If $R_1$ does not lie wholly in $S(P, \epsilon)$, there is a point, $P_1$, in $R_1$, such that $\rho(P, P_1)=\epsilon$.

In a similar manner, we can show the existence, in $S(P, \epsilon/2)$, of arcs $C_1x_1D_1$ and $C_2y_2D_2$, where $<C_1x_1D_1>$ and $<C_2y_2D_2>$ lie in $K_1$ and $K_2$, respectively, and the end points of these arcs lie on $A_2PB_2$ in the order $A_2C_1C_2PD_2D_1B_2$. Denote the simple closed curve formed by these two arcs, together with the portions $C_1C_2$ and $D_1D_2$ of $A_2PB_2$, by $J_2$. Denote the component of $M-J_2$ that contains $P$ by $R_2$.

By VII, $K_1-<C_1x_1D_1>$ is the sum of two components, one of which, $N_1$, is bounded by the simple closed curve $F_1=C_1x_1D_1+C_1PD_1$; similarly, that component of $K_2-<C_2y_2D_2>$ bounded by the curve $F_2=C_2y_2D_2+C_2PD_2$ denote by $N_2$. Then we can show, as above, that

$$R_2 = N_1 + N_2 + < C_2PD_2 > .$$

Consequently, $R_2$ is a subset of $R_1$, and does not contain the point $\alpha$ defined above.

We may continue in this way indefinitely, obtaining simple closed curves $J_3, J_4, J_5, \cdots$, with corresponding components $R_3, R_4, R_5, \cdots$, all containing $P$, but not $\alpha$, and each component a subset of the preceding; also, requiring that for each positive integer $n$, the curve $J_n$ lie in $S(P, \epsilon/n)$.

For some $n$, $R_n$ is a subset of $S(P, \epsilon)$. For if not, there is for every $n$ a point $P_n$ of $R_n$ such that $\rho(P, P_n)=\epsilon$. Since $M$ is compact, the set of points $\sum_{n=1}^{\infty} P_n$ has at least one limit point $Q$, which is in every set $R_n$ and such that $\rho(P, Q)=\epsilon$. Common to all the sets $R_1, R_2, R_3, \cdots$, there is a continuum $C$, which contains $P$ and $Q$.

By VI, $M-P$ is connected, and therefore contains an arc $t$ with end

points $Q$ and $\alpha$.* There exists a positive integer $k$ such that $\epsilon/k$ is less than the distance between $P$ and the arc $t$. Then the set $C+t$ is a connected subset of $M-J_k$, since $C$ lies in $R_k$ and $t$ cannot meet $J_k$. However, $P$ lies in $R_k$ and $\alpha$ lies in $M-(J_k+R_k)$. Thus the supposition that for no $n$ does $R_n$ lie in $S(P, \epsilon)$ leads to a contradiction.

The conclusion of Theorem 2 now follows in either one of the following ways: (1) by virtue of the Moore-Kline-Gawehn† definition of the simple closed surface, whose conditions are now seen to be fulfilled; or (2) by virtue of R. L. Moore's Axioms $\Sigma_1$ for plane analysis situs.‡ For it is easy to see, in view of what has been shown above, that if $Q$ is an arbitrary point of $M$, then $M-Q$ is a topological plane, in that it satisfies the axioms $\Sigma_1$. Thus, if $J$ is a simple closed curve of $M-Q$, let that component of $M-J$ which does not contain $Q$ be called a region.

3.§ We shall prove the following theorem:

THEOREM 3 (Converse of the Jordan-Brouwer Separation Theorem in $E_3$). *Let $K$ be a closed and bounded set in $E_3$, such that $E_3-K=S_1+S_2$, where $S_1$ and $S_2$ are mutually exclusive and*

(1) *every arc from a point of $S_1$ to a point of $S_2$ contains at least one point of $K$;*

(2) *if $P$ is a point of $K$ and $Q$ a point not in $K$, then in every neighborhood of $P$ there is a point $P'$ such that there is an arc from $P'$ to $Q$ lying except for possibly $P'$ wholly in $E_3-K$;*

(3) *the 0-chains of $S_i (i=1, 2)$ are uniformly homologous to zero in $S_i$; and*

(4) *the Betti number* (mod 2) $R^1(E_3-K)=1$.

*Then $K$ is a simple closed surface.*

We shall show that the set $K$ satisfies all the conditions of Theorem 2.

I. *The set $K$ is connected.* For if not, it is the sum of two mutually separated sets, $K_1$ and $K_2$. Let $P_i$ be a point of $K_i$ $(i=1, 2)$. By a theorem of Knaster and Kuratowski,¶ there exists a continuum $C$ in $E_3-K$ which separates $P_1$ and $P_2$. The continuum $C$ lies wholly in one of the sets $S_1$, $S_2$.

* Cf. R. L. Moore, Mathematische Zeitschrift, loc. cit.

† Cf. Bulletin abstract of Moore and Kline, loc. cit., and I. Gawehn, loc. cit.

‡ R. L. Moore, *On the foundations of plane analysis situs*, these Transactions, vol. 17 (1916), pp. 131–164. See also R. L. Wilder, *Concerning R. L. Moore's axioms $\Sigma_1$ for plane analysis situs*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 752–760.

§ The contents of this section were presented to the Society August 30, 1929.

¶ B. Knaster and C. Kuratowski, *Sur les ensembles connexes*, Fundamenta Mathematicae, vol. 2 (1921), pp. 206–255. The theorem referred to here is Theorem 37, p. 233. This theorem holds true in $E_n$, since the theorem of Brouwer used in its proof holds in $E_n$, for both bounded and unbounded sets.

For if it contains points of both $S_1$ and $S_2$, then $S_1$, say, contains a limit point of $S_2$. But then, since $K$ is closed, this limit point would be joined to a point of $S_2$ by an arc that does not meet $K$, a violation of condition (1). Then $C$ lies in $S_1$, say.

Now in the neighborhood of any point $P$ of $K$ there lie points of both $S_1$ and $S_2$. For if we let $U$ be any neighborhood of $P$ and $x$ a point of $S_1$, say, there is, by condition (2), a point $P'$ in $U$ which is joined to $x$ by an arc $P'x$ and such that $P'x-P'$ lies in $E_3-K$. It is clear that $P'x-P'$ must lie in $S_1$, since $x$ is in $S_1$; and as $U$ must contain points of $P'x-P'$, there is a point of $S_1$ in $U$.

Accordingly, in the same domain complementary to $C$ that contains $P_1$ there is a point $x_1$ of $S_2$. By condition (2) there is, in the same domain complementary to $C$ as $P_2$, a point $x_2$ such that there is an arc $x_1x_2$ which lies, except possibly for $x_2$, wholly in $E_3-K$. Then $x_1x_2-x_2$ lies in $S_2$, and there is no point of $C$ on $x_1x_2$, since $C$ is in $S_1$. But $x_1$ and $x_2$ must be separated by $C$ in $E_3$. Thus the supposition that $K$ is not connected leads to a contradiction.

II. *The set $K$ is connected im kleinen.* For if not, it follows from a theorem of R. L. Moore* that there exist two concentric spheres, $R_1$ and $R_2$, and a sequence of subcontinua of $K$, namely $M_\infty$, $M_1$, $M_2$, $M_3$, $\cdots$, such that (1) each of these continua contains at least one point of $R_1$ and $R_2$, respectively, but no point exterior to $R_1$ or interior to $R_2$, (2) no two of these subcontinua have a point in common, and no two of them contain points of any connected subset of $K$ that lies wholly in $R_1+R_2+I$, where $I$ is the annular domain bounded by $R_1$ and $R_2$, (3) $M_\infty$ is the sequential limiting set of the sequence of continua $M_1$, $M_2$, $M_3$, $\cdots$.

Let $P$ be a point of $M_\infty$ in $I$, and let $R$ be a spherical neighborhood of $P$ such that $R'$ lies in $I$ (where $R'$ denotes $R$ together with its boundary). Let us consider $R$ as a space, $\overline{E}_3$, and if $M$ is a point set, let us denote the product $M \cdot R$ by $\overline{M}$.

There exists a positive integer $m$ such that $\overline{M}_i$, for $i \geqq m$, is non-vacuous. The set $\overline{K}$ is closed relative to $\overline{E}_3$, and by condition (2) of Moore's theorem just quoted, no connected subset of $\overline{K}$ joins $\overline{M}_i$ and $\overline{M}_j$ $(i \neq j;\ i, j \geqq m)$. There exists a spherical neighborhood, $U$, of $P$ in $\overline{E}_3$, such that any 0-chain of $S_i$ in $U$ is homologous to zero in $\overline{S}_i$. Let $i$ and $j$ be values such that $\overline{M}_i$

* R. L. Moore, *A characterization of Jordan regions by properties having no reference to their boundaries*, Proceedings of the National Academy of Sciences, vol. 4 (1918), pp. 364–370. The theorem referred to here is not given any explicit statement, in theorem form, in this paper, but will be found in such form in the same author's *Report on continuous curves from the viewpoint of analysis situs*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 296–297.

and $\overline{M}_j$ have points $P_i$ and $P_j$, respectively, in $U$. There exists a continuum, $C$ (continuum relative to $\overline{E}_3$) that lies in $\overline{E}_3 - \overline{K}$ and separates $P_i$ and $P_j$ in $\overline{E}_3$. For there is a separation of $\overline{K}$ into two mutually separated sets containing $\overline{M}_i$ and $\overline{M}_j$, respectively,* and the theorem of Knaster and Kuratowski used in I applies in $\overline{E}_3$.

Now, as shown in a similar case in I, $C$ must lie wholly in, say, $S_1$. But in every neighborhood of $P_i$, and of $P_j$, there are points of $S_2$; in particular, there are such points in neighborhoods of $P_i$ and $P_j$ that lie in $U - C \cdot U$, and these bound 1-chains of $\overline{S}_2$. As such chains must contain points of $C$, and hence of $S_1$, a contradiction is established, and $K$ must be connected im kleinen.

III. *The sets $S_1$ and $S_2$ are connected.* Consider $S_1$, and let $x$ and $y$ be any two of its points. Let $P$ be any point of $K$, and let $U_1$ be a spherical neighborhood of $P$. By condition (3) there is a spherical neighborhood $U_2$ of $P$ such that if $a^0$ and $b^0$ are two 0-cells of $S_1$ in $U_2$, then $a^0$ and $b^0$ bound a 1-chain of $S_1$ in $U_1$. Now, as already shown, $U_2$ contains two points, $x_1$ and $y_1$, which are joined to $x$ and $y$, respectively, by arcs $xx_1$ and $yy_1$ in $S_1$. There is no loss of generality in assuming that $x_1$ and $y_1$ are 0-cells of the subdivision of $E_3$, and accordingly bound a 1-chain $L^1$ of $S_1$ in $U_1$. Then $xx_1 + L^1 + yy_1$ is a connected subset of $S_1$ joining $x$ and $y$, and therefore $S_1$ is connected.

IV. *No arc of $K$ disconnects $K$.* For let $t$ be an arc of $K$ and suppose that $K - t = K_1 + K_2$, where $K_1$ and $K_2$ are mutually separated sets. The sets $K_1 + t$ and $K_2 + t$ are closed.

Let $P$ and $Q$ be points of $S_1$ and $S_2$, respectively, which are also 0-cells of the subdivision of $E_3$. By condition (2), $K_1 + t$ does not separate $P$ and $Q$ in $E_3$; let $L_1^1$ be a 1-chain bounded by $P$ and $Q$ in the complement of $K_1 + t$. Similarly, let $L_2^1$ be a 1-chain bounded by $P$ and $Q$ in the complement of $K_2 + t$. Since the closed 1-chain $L_1^1 + L_2^1$ cannot link the arc $t$,† it follows from a theorem of J. W. Alexander† that $P + Q$ bounds a 1-chain in $E_3 - K$; in other words, condition (1) is violated. Hence $t$ does not separate $K$.

V. *If $J$ is a simple closed curve of $K$, then $K - J$ is not connected.* Suppose that $K - J$ is connected.

By the Alexander duality relation (see introduction) there is a closed 1-chain $\Gamma^1$ which links $J$. We shall first show that $\Gamma^1$ has points in both $S_1$ and $S_2$.

Suppose $\Gamma^1$ is contained in $S_1 + K$. Then it is clear that $\Gamma^1$ contains points of $K$, for if not, by condition (4) we should have

* This follows from the fact that the closed set $K \cdot R'$ is separated between $M_i \cdot R'$ and $M_j \cdot R'$.

† J. W. Alexander, these Transactions, loc. cit.

$$\Gamma^1 \sim 0 \qquad (\text{mod } 2,\ E_3 - J),$$

which is impossible since $\Gamma^1$ links $J$. Let the distance between $J$ and $\Gamma^1$, as ordinarily defined for point sets, be denoted by $\epsilon$. By condition (3) of the theorem, there is a number $\delta_\epsilon$ such that if $\Gamma^0$ is a closed 0-chain of $S_1$ of diameter less than $\delta_\epsilon$, then there is a 1-chain $\Gamma_1^1$ such that

$$\Gamma_1^1 \equiv \Gamma^0 \qquad (\text{mod } 2,\ S_1)$$

where the diameter of $\Gamma_1^1$ is less than $\frac{1}{4}\epsilon$.

We may assume, without loss of generality, that the 1-chain $\Gamma^1$ is irreducible, and that its 1-cells have a given cyclic order. We may also assume that the one-cells of $\Gamma^1$ are all of diameter less than $\frac{1}{4}\delta_\epsilon$.

The cells of $\Gamma^1$ may be divided into two classes, according to whether their boundaries lie wholly in $S_1$ or not. We may start with a given cell $c^1$, and consider the successive cells of $\Gamma^1$ in their cyclic order. If both end-cells of $c^1$ lie in $S_1$, then $c^1$ may be replaced by a 1-chain $K^1$ which has the same boundary, but which lies entirely in $S_1$ and has a diameter less than $\frac{1}{4}\epsilon$. We may agree to say that the cell $c^1$ has been transformed into the *chain-cell* $K^1$. Proceeding to the next 1-cell of $\Gamma^1$, which we may denote by $d^1$, suppose that one of its end-cells is on $K$. Denote the cell that it has in common with $c^1$ by $a^0$, and the cell on $K$ by $b^0$. There is a 0-cell, $\bar{b}^0$, in $S_1$, whose distance from $b^0$ is less than $\frac{1}{4}\delta_\epsilon$, and since the distance from $a^0$ to $\bar{b}^0$ is less than $\delta_\epsilon$, these two cells bound a 1-chain $K_1^1$ in $S_1$, whose diameter is less than $\frac{1}{4}\epsilon$. We shall replace $b^0$ by $\bar{b}^0$, and $d^1$ by $K_1^1$, and say that $b^0$ and $d^1$ have been transformed, respectively, into $\bar{b}^0$ and $K_1^1$. Suppose the next cell on $\Gamma^1$ is $e^1$, and that its end-cell distinct from $b^0$, viz. $c^0$, is also on $K$. There is a 0-cell $\bar{c}^0$ in $S_1$ whose distance from $c^0$ is less than $\frac{1}{4}\delta_\epsilon$. Clearly the distance from $\bar{b}^0$ to $\bar{c}^0$ is less than $\delta_\epsilon$, and hence these two cells bound a 1-chain $K_2^1$ in $S_1$ of diameter less than $\frac{1}{4}\epsilon$. We shall let $\bar{c}^0$ and $K_2^1$ be the transforms, respectively, of $c^0$ and $e^1$.

Proceeding through all of the cells of $\Gamma^1$ as just indicated, we replace each cell of $\Gamma^1$ by its transform, and obtain a new closed 1-chain, $\Gamma_1^1$, which we may call the transform of $\Gamma^1$. Now $\Gamma_1^1$ lies wholly in $S_1$, and accordingly, by application of condition (4),

$$\text{(a)} \qquad \Gamma_1^1 \sim 0 \qquad (\text{mod } 2,\ E_3 - J).$$

We shall show that relation (a) is impossible.

We note that if $c^1$ is any 1-cell of $\Gamma^1$, then every point of the transform, $K^1$, of $c^1$, is at a distance less than $\frac{1}{2}\epsilon$ from either end-cell of $c^1$. That is, both of the chains $c^1$ and $K^1$ lie within a spherical neighborhood $T$ of radius $\epsilon$ which has its center at one end-cell of $c^1$ and accordingly contains no point of

$J$. By introducing, if necessary, two new 1-chains, each bounded by an end-cell of $c^1$ and its transform, a closed 1-chain $L^1$ is obtained in $T$ which contains $c^1$ and $K^1$ and such that

(b) $$L^1 \sim 0 \qquad (\text{mod } 2, T, E_3 - J).$$

By adding homology (a) and all homologies of type (b), we have

(c) $$\Gamma_1^1 + \sum L^1 \sim 0 \qquad (\text{mod } 2, E_3 - J).$$

However, we have that, adding modulo 2,

(d) $$\Gamma_1^1 + \sum L^1 = \Gamma^1$$

(it can be assumed that only one chain is introduced between each 0-cell of $\Gamma^1$ and its transform). Combining relations (c) and (d) we have

$$\Gamma^1 \sim 0 \qquad (\text{mod } 2, E_3 - J).$$

This contradicts the fact that $\Gamma^1$ links $J$. Hence $\Gamma^1$ must have points in both $S_1$ and $S_2$.

We shall now show that the supposition that $K-J$ is connected is inconsistent with the fact that $\Gamma^1$ links $J$, and has points in both $S_1$ and $S_2$. It will first be necessary to make a transformation of $\Gamma^1$.

Let $\epsilon_1$ be a positive number less than $\frac{1}{4}\epsilon$. Since $K$ is uniformly connected im kleinen (being a bounded, connected im kleinen continuum), there is a positive number $\epsilon_2$, such that if $P$ and $Q$ are points of $K$ at a distance apart less than $\epsilon_2$, then $P$ and $Q$ are joined by an arc of $K$ of diameter less than $\epsilon_1$. Let $\delta_{\epsilon_2}$ be a positive number such that if $P$ and $Q$ are 0-cells of $S_i$ ($i=1$ or 2) at a distance apart less than $\delta_{\epsilon_2}$, then $P$ and $Q$ bound a 1-chain in $S_i$ of diameter less than $\frac{1}{4}\epsilon_2$. We may assume that the 1-cells of $\Gamma^1$ are all of diameter less than $\frac{1}{4}\delta_{\epsilon_2}$.

The transformation that we now effect on $\Gamma^1$ is very similar to the one described above, except that the new constants just defined are involved. Supposing that we start with a cell $c^1$ of $\Gamma^1$ whose end-cells are both in $S_1$, we replace it by a chain-cell $K^1$ with the same boundary and diameter less than $\frac{1}{4}\epsilon_2$. If the next cell $b^1$ has an end-cell on $K$, we transform it into a chain-cell in $S_1$. Of course we ultimately come to a cell $c^1$ one of whose end-cells (the one last affected by a transformation), $a^0$, is in $S_1+K$, and the other, $b^0$, is in $S_2$. (Indeed, if $\Gamma^1$ had none of its bounding 0-cells in $S_2$, the whole chain $\Gamma^1$ could be transformed into a chain in $S_1$, just as in the above proof, and a contradiction obtained as before. Similarly, if we encounter a case of one bounding 0-cell of $\Gamma^1$ in $S_2$ ($S_1$) and the neighboring 0-cells in $S_1$ ($S_2$), the 1-cells which they bound may be transformed into chain-cells lying in

$S_1$ ($S_2$).) It is unnecessary to use condition (3) here, since any 1-chain $K^1$ of diameter less than $\frac{1}{4}\epsilon_2$ and bounded by $\bar{a}^0$ and $b^0$ will do for the transform of $c^1$ ($\bar{a}^0$ being the transform of $a^0$). The next 0-cell, $c^0$, is in $S_2$ (else we would have proceeded as indicated in the above parenthesis) and the transformation to be effected is obvious. The subsequent transformations should also be obvious, since we now proceed from $S_2$ in just such a manner as we proceeded from $S_1$, viz., by not actually "crossing" into $S_1$ until two successive bounding 0-cells are encountered in $S_1$.

The outcome of these transformations on the cells of $\Gamma^1$ is to obtain a new 1-chain, which we shall still denote by $\Gamma^1$ and which still links $J$ (this can be shown as above), but which has the following properties: (1) its intersections with $K$ occur on certain chain-cells, $K_1^1, K_2^1, \cdots, K_n^1$ in the order named; we shall henceforth call these the *crossing-cells* of $\Gamma^1$; (2) if $K_i^1$ is any crossing-cell, the diameter of $K_i^1$ is less than $\frac{1}{4}\epsilon_2$; (3) the boundary cells of $K_i^1$ lie in $S_1$ and $S_2$, respectively; (4) if $K_i^1$ and $K_{i+1}^1$ are bounded by the cells $a^0$, $b^0$, $c^0$ and $d^0$ (these occurring in the order named), then the cells $a^0$ and $d^0$ are both in $S_1(S_2)$; in other words, if one crossing-cell leads from $S_1$ to $S_2$, the next leads from $S_2$ to $S_1$. The last property is very important, and shows that there is an even number of crossing-cells.

We shall now proceed to replace $\Gamma^1$ by a 1-chain which still links $J$ but has only *two* crossing-cells.

Starting with $K_1^1$, suppose that the bounding 0-cells of $K_1^1$ and $K_2^1$ occur in the order $a^0$, $b^0$, $c^0$, $d^0$, where $a^0$ is in $S_1$. In $S_1$, let $A_1^1$ be a new 1-chain bounded by $a^0$ and $d^0$, and in $S_2$ let $A_2^1$ be the 1-chain of $\Gamma^1$ which lies in $S_2$ and is bounded by $b^0$ and $c^0$. Denote the portion of $\Gamma^1$ from $d^0$ to $a^0$ by $B^1$. Now we cannot have both of the following relations:

$$A_1^1 + K_1^1 + A_2^1 + K_2^1 \sim 0, \tag{e}$$

$$A_1^1 + B^1 \sim 0 \pmod{2, E_3 - J}, \tag{f}$$

since the sum of these homologies would imply the homology

$$\Gamma^1 \sim 0 \pmod{2, E_3 - J}.$$

If (e) fails to hold, we have secured the type of 1-chain linking $J$ that we set out to obtain, viz., a chain with only two crossing-cells. If (e) holds, we can proceed with the 1-chain

$$\Gamma^1 + K_1^1 + A_2^1 + K_2^1 + A_1^1 = A_1^1 + B^1 \pmod{2}$$

which has two less crossing-cells than $\Gamma^1$, viz., $K_3^1, \cdots, K_n^1$, by next considering the crossing-cells $K_3^1$ and $K_4^1$.

In any case, we eventually arrive at a closed 1-chain which links $J$, and

has only two of the original crossing-cells of $\Gamma^1$. We shall continue to call this chain $\Gamma^1$, and we shall suppose its crossing-cells are $K_1^1$ and $K_2^1$, their boundary cells being denoted, as before, by $a^0$, $b^0$, $c^0$, $d^0$. Denote the portion of $\Gamma^1$ lying in $S_1$ and bounded by $a^0+d^0$ by $L_1^1$ and the portion in $S_2$ bounded by $b^0+c^0$ by $L_2^1$.

Because of the way in which the above transformation was carried out, at least one of the boundary cells of $K_i^1$ $(i=1, 2)$ is identical with a cell of the original chain $\Gamma^1$, so that a spherical neighborhood $T_i$ of diameter $\epsilon_2$ about $K_i^1$ certainly encloses no points of $J$. We may assume that $K_1^1$ and $K_2^1$ are homeomorphic with simple arcs, and on $K_i^1$ $(i=1, 2)$ let $x_i$ and $y_i$ be points of $K$ such that the portions $a^0x_1$ $(d^0x_2)$ and $b^0y_1$ $(c^0y_2)$ of $K_1^1$ $(K_2^1)$ contain no points of $K$ other than $x_1$ and $y_1$ ($x_2$ and $y_2$).

Since $x_1$ and $x_2$ lie in $K-J$, and $K-J$ is connected, there is an arc $x_1x_2$ in $K-J$.* Denote the distance between $x_1x_2$ and $J$ by $\epsilon_3$, let $\epsilon_4$ denote a positive number less than both $\epsilon_1$ and $\epsilon_3$, and let $\delta_{\epsilon_4}$ be a positive number such that if two 0-cells of $S_i$ are at a distance apart less than $\delta_{\epsilon_4}$, they bound a 1-chain in $S_i$ of diameter less than $\frac{1}{4}\epsilon_4$ (condition (3)).

Let $F_1, F_2, \cdots, F_m$ be points on the arc $x_1x_2$, where $F_1=x_1$ and $F_m=x_2$, occurring in the order named, and so selected that

$$\delta(F_iF_{i+1}) < \tfrac{1}{4}\delta_{\epsilon_4} \qquad (i = 1, 2, \cdots, m-1),$$

where $F_iF_{i+1}$ denotes the portion of $x_1x_2$ between $F_i$ and $F_{i+1}$. Let $P_1, P_2, \cdots, P_m$ be a set of 0-cells in $S_1$ such that $P_1$ is on the portion $a^0x_1$ of $K_1^1$ and $P_m$ on the portion $d^0x_2$ of $K_2^1$, and the following relations hold:

$$\rho(P_i, F_i) < \tfrac{1}{4}\epsilon_4,$$
$$\rho(P_i, P_{i+1}) < \delta_{\epsilon_4}.$$

According to the definition of $\delta_{\epsilon_4}$, there exists for each $i$ an open 1-chain $H_i^1$ in $S_1$ bounded by $P_i$ and $P_{i+1}$, and such that $\delta(H_i^1)<\frac{1}{4}\epsilon_4$. We define $C_1^1=\sum_{i=1}^{m-1} H_i^1$, and we also note that every point of $H_i^1$ lies within a distance $\frac{1}{2}\epsilon_4$ of $F_i$.

Let $T_i^1$ $(i=1, 2)$ be a spherical neighborhood concentric with $T_i$ and of diameter $2\epsilon_1$. Clearly $T_i^1$ contains no point of $J$. According to the definition of $\epsilon_2$, and because of the fact that the distance from $x_i$ to $y_i$ is less than $\epsilon_2$ (the diameter of $K_i^1$ being less than $\epsilon_2$), there is an arc $t_i$ in $K$ which lies wholly in $T_i$ and has end points $x_i$ and $y_i$. We can approximate the arc $t_1$ by a $\frac{1}{4}\delta_{\epsilon_4}$-chain† of 0-cells in $S_2$, with first point, $z_1$, on the portion $y_1b^0$ of $K_1^1$, and

* R. L. Moore, Mathematische Zeitschrift, loc. cit.

† The word "chain" is here used in the ordinary point-set theoretic sense; i.e., if $\epsilon$ is a positive number, and $P$ and $Q$ are points, an $\epsilon$-chain from $P$ to $Q$ is a set of points $P_1, P_2, \cdots, P_n$, where $P_1=P$ and $P_n=Q$, and $\rho(P_i, P_{i+1})<\epsilon$ $(i=1, 2, \cdots, n-1)$.

last point, $z_h$, and such that not only is every point of the chain at a distance less than $\frac{1}{4}\epsilon_4$ from some point of $t_1$, but in particular $\rho(z_h, F_1) < \frac{1}{4}\epsilon_4$. Then, in a manner similar to that used in obtaining $C_1^1$, we obtain a 1-chain $B_1^1$ bounded by $z_1+z_h$, lying wholly in $S_2$, and such that every point of $B_1^1$ is within a distance $\frac{1}{2}\epsilon_4$ from some point of $t_1$. It is clear that a spherical neighborhood $T_1^2$ concentric with $T_1$ and of diameter $3\epsilon_1$, will contain $B_1^1$, and enclose no points of $J$.

We can obtain similarly a 1-chain $B_2^1$ with reference to $t_2$, so that the boundary of $B_2^1$ is a point $w_1$ on the portion $y_2c^0$ of $K_2^1$ and a point $w_k$ in $S_2$ whose distance from $F_m$ is $< \frac{1}{4}\epsilon_4$. Let $T_2^2$ be a spherical neighborhood analogous to $T_1^2$.

Let $Q_1, Q_2, \cdots, Q_m$ be points in $S_2$ chosen with reference to the points $\{F_i\}$ just as the points $\{P_i\}$ were chosen in $S_1$, and so that $Q_1 = z_h$ and $Q_m = w_k$. On these points can be built up a 1-chain $B_3^1$ of $S_2$, bounded by $Q_1$ and $Q_m$, and approximating the arc $x_1x_2$ in a manner similar to that in which $C_1^1$ approximates $x_1x_2$.

We now define the following 1-chains:

$$
\begin{aligned}
C_2^1 &= \sum_{i=1}^{3} B_i^1, \\
\Gamma_1^1 &= L_1^1 + a^0P_1 + C_1^1 + P_md^0, \\
\Gamma_2^1 &= C_1^1 + P_1z_1 + C_2^1 + w_1P_m, \\
\Gamma_3^1 &= C_2^1 + z_1b^0 + L_2^1 + c^0w_1,
\end{aligned}
$$

where $a^0P_1$, $P_md^0$, etc., are portions of $K_1^1$ and $K_2^1$. We note that

$$
\Gamma^1 = \sum_{i=1}^{3} \Gamma_i^1 \pmod 2. \tag{g}
$$

The following homologies follow from condition (4) of the theorem:

$$
\Gamma_1^1 \sim 0, \tag{h}
$$

$$
\Gamma_3^1 \sim 0 \pmod{2, E_3 - J}. \tag{i}
$$

Consequently, since $\Gamma^1$ links $J$, it follows from relations (g), (h) and (i) that $\Gamma_2^1$ links $J$.

If we join $P_1$ and $z_h$ by a 1-chain $B_4^1$ every point of which is at a distance less than $\frac{1}{4}\epsilon_4$ from $x_1$ $(=F_1)$ and $P_m$ and $w_k$ similarly by a 1-chain $B_5^1$, it is easy to see that since the following relations hold:

$$
B_1^1 + B_4^1 + P_1z_1 \sim 0 \qquad (T_2^1, E_3 - J),
$$

$$
B_2^1 + B_5^1 + P_mw_1 \sim 0 \qquad (T_2^2, E_3 - J),
$$

the closed chain

$$\Gamma = C_1^1 + \sum_{i=3}^{5} B_i^1$$

links $J$.

However, this is impossible for the following reasons. The points $P_i$ and $Q_i$ $(i=2, 3, \cdots, m-1)$ bound a 1-chain $N_i^1$, every point of which is at a distance less than $\frac{1}{4}\epsilon_4$ from $F_i$. For convenience of notation, let $N_1^1 = B_1^1$ and $N_m^1 = B_5^1$. Let $U_i$ $(i=1, 2, \cdots, m)$ be a spherical neighborhood of $F_i$ of radius $\epsilon_4$. Then every point of the closed chain

$$E_i^1 = H_i^1 + N_i^1 + \overline{H}_i^1 + N_{i+1}^1,$$

where $\overline{H}_i^1$ is the portion of $B_3^1$ from $Q_i$ to $Q_{i+1}$, lies in $U_i$. Hence

$$\text{(j)} \qquad E_i^1 \sim 0 \qquad (\text{mod } 2, E_3 - J).$$

But

$$\Gamma = \sum_{i=1}^{m} E_i^1 \qquad (\text{mod } 2),$$

and by adding all homologies (j), we get

$$\Gamma \sim 0 \qquad (\text{mod } 2, E_3 - J).$$

Thus, on the assumption that $K-J$ is connected, we are able to find a closed chain $\Gamma$, homologous to the chain $\Gamma^1$ which links $J$, but which does not itself link $J$. As this is impossible, $K-J$ cannot be connected.

VI. *If $J$ is a simple closed curve of $K$, then $K-J$ contains at most two components.* For suppose $K-J$ contains at least 3 components, $M_1$, $M_2$, and $M_3$. Since $K$ is a continuous curve, each component of $K-J$ is also a quasi-component of $K-J$,* and it is therefore easy to see, with the result of section IV, that each component of $K-J$ has all of $J$ as its boundary.

There exists on $J$ a set of points occurring in the following order, $P_1A_1Q_1B_1P_2B_2Q_2A_2P_1$, and such that there is an arc $A_ix_iB_i$ $(i=1, 2)$ lying, except for $A_i$ and $B_i$, wholly in $M_i$. Define a simple closed curve $J_1$ as follows:

$$J_1 = \sum_{i=1}^{2} A_ix_iB_i + A_1P_1A_2 + B_1P_2B_2,$$

where $A_1P_1A_2$ and $B_1P_2B_2$ are arcs of $J$. Then the set $K-J_1$ is connected, in contradiction to the result of section V. For if we denote the set

* Cf. R. L. Wilder, *A characterization of continuous curves by a property of their open subsets*, Fundamenta Mathematicae, vol. 11 (1928), pp. 127–131.

$K-(M_1+M_2+J)$ by $R$, it is clear that the set $R_1=R+\sum_{i=1}^{2}<A_iQ_iB_i>$ is connected. Also, the set $M_1-<A_1x_1B_1>$ contains only components that have boundary points on either $<A_1Q_1B_1>$ or $<A_2Q_2B_2>$. For suppose $M_1-<A_1x_1B_1>$ contains a component $C$ with boundary points only on the set $t=A_2P_1A_1x_1B_1P_2B_2$. Then the set

$$K-t=C+\left[(M_1-<A_1x_1B_1>-C)+M_2+R+\sum_{i=1}^{2}<A_iQ_iB_i>\right]$$

is not connected, contrary to the result of IV. In a like manner it is shown that all components of $M_2-A_2x_2B_2$ have limit points on $\sum_{i=1}^{2}<A_iQ_iB_i>$, and it is clear that the set

$$K-J_1=R_1+\sum_{i=1}^{2}(M_i-A_ix_iB_i)$$

is connected.

VII. *Denoting the two components of $K-J$ by $M_1$ and $M_2$, respectively, the sets $M_1$ and $M_2$ are uniformly connected im kleinen.* For suppose $M_1$ is not uniformly connected im kleinen. Then there is a point $P$ on $J$, a spherical neighborhood $R$ of $P$, and a sequence of pairs of points of $M_1$, $\{P_n, Q_n\}$, such that $\lim_{n\to\infty}P_n=P$ and $\lim_{n\to\infty}Q_n=P$, and for no $n$ are $P_n$ and $Q_n$ joined in $R$ by any connected subset of $M_1$. Denote the set $M_i\cdot R$ by $\overline{M}_i$ $(i=1, 2)$, and the frontier of $R$ by $F$.

Denote by $C$ the component of $K\cdot R$ determined by $P$. Let $R_1$ be a spherical neighborhood of $P$ of smaller radius than $R$, which encloses no points of $J$ that are not on the arc, $t$, component of $J\cdot R$ determined by $P$. We shall denote $J\cdot R$ by $t_1$.

Let $R_2$ be another spherical neighborhood of $P$ such that any point of $K$ within $R_2$ is joined to $P$ by an arc of $K$ that lies in $R_1$ and such that $R_2$ encloses no points of $K-C$. Let $P_k$, $Q_k$ be a pair of points of the sequence $\{P_n, Q_n\}$ lying in $R_2$, and let the component of $\overline{M}_1$ determined by $P_k$ be denoted by $H_1$; let $\overline{M}_1-H_1=H_2$. Then $H_1$ and $H_2$ are mutually separated.

In $R_2$, let $x_1$ and $x_2$ be points of $S_1$ and $S_2$, respectively, that are 0-cells of the subdivision of $E_3$.

(a) The set $C$ separates $R$ between $x_1$ and $x_2$. For $x_1$ and $x_2$ bound a chain $\Gamma_1^1$ in $E_3-[F+(K-C)]$, and if they bound a chain $\Gamma_2^1$ in $E_3-(F+C)$, then, since the closed chain $\Gamma_1^1+\Gamma_2^1$ is homologous to zero in $R$, it cannot link the product $[F+(K-C)]\cdot[F+C]$, which is a subset of $F$, and accordingly $x_1+x_2$ is homologous to zero in $E_3-K$ (Alexander's Corollary $W_i$, loc. cit.). This is a contradiction of condition (1) of the theorem.

(b) Under the assumption that $M_1$ is not uniformly connected im kleinen

in the neighborhood of $P$, $C$ does not separate $x_1$ and $x_2$ in $R$. Let $y_1$ and $y_2$ be the first points of $K$ on a straight line from $x_1$ to $x_2$, in the orders $x_1$ to $x_2$ and $x_2$ to $x_1$, respectively. There exist arcs $y_1P_k$, $y_1Q_k$, $y_2P_k$ and $y_2Q_k$ of $K$ in $R_1$. By the use of the usual approximation based on condition (3) of the theorem, we obtain a closed 1-chain, $\Gamma^1$, consisting of two open chains $\Gamma_1^1$ and $\Gamma_2^1$ bounded by $x_1$ and $x_2$, such that $\Gamma^1$ lies in $R_1$, and cuts $K$ only in certain arbitrarily small neighborhoods of $P_k$ and $P_j$, respectively. The following congruences hold:

$$\Gamma_1^1 \equiv x_1 + x_2 \qquad [\text{mod } 2, E_3 - (F + H_1')],$$
$$\Gamma_2^1 \equiv x_1 + x_2 \qquad [\text{mod } 2, E_3 - (F + H_2' + \overline{M}_2')]$$

(where $H_1'$ denotes $H_1$ together with all its limit points, etc.). We may assume that the product of $H_1'$ and $H_2'$ is a subset of $t_1'$, so that if we can show that $\Gamma^1$ bounds in $E_3-(F+H_1')\cdot(F+H_2'+\overline{M}_2')$ or, since the latter set contains $E_3-(F+t_1)$, if we can show that $\Gamma^1$ bounds in $E_3-(F+t_1)$, we shall have that $x_1$ and $x_2$ are not separated in $R$ by $C$ (Alexander's Corollary $W_i$, loc. cit.).

Since $M_1$ is connected, it can be shown, by the methods employed in section V, that $\Gamma^1$ does not link $J$. Consequently, there exists a chain $\Gamma_1^2$ such that

$$\Gamma_1^2 \equiv \Gamma^1 \qquad (\text{mod } 2, E_3 - J, E_3 - t').$$

But since $\Gamma^1$ lies in $R_1$, and the latter contains no points of $F+J-t$, we have

$$\Gamma_2^2 \equiv \Gamma^1 \qquad [\text{mod } 2, E_3 - (F + J - t)].$$

Now since the product of $F+J-t$ and $t'$ is just two points, and neither $\Gamma_1^2$ nor $\Gamma_2^2$ meets the arc $J-t$ joining these points, we have that

$$\Gamma_1^2 + \Gamma_2^2 \sim 0 \quad [\text{mod } 2, E_3 - t'\cdot(F + J - t)].$$

Hence, by Alexander's Corollary $W_i$ (loc. cit.),

$$\Gamma^1 \sim 0 \qquad [\text{mod } 2, E_3 - (F + t_1)].$$

As this is the relation we wished to prove, in order to show that the 0-chain $x_1+x_2$ bounds in $R-C$, we have shown that $x_1$ and $x_2$ are not separated by $C$ in $R$.

As (a) and (b) are in contradiction, the assumption that $M_1$ is not uniformly connected im kleinen cannot hold. A like statement of course holds for $M_2$, and the theorem is proved.

The following theorem follows immediately from Theorem 3, from which it is only slightly different:

THEOREM 4 (Converse of the Jordan-Brouwer Separation Theorem in $E_3$). *Let $K$ be a closed and bounded set in $E_3$, such that* (1) *the Betti numbers* (mod 2) $R^0(E_3-K)$ *and* $R^1(E_3-K)$ *are respectively equal to* 2 *and* 1, (2) *if $D$ is a component of $E_3-K$, the* 0*-chains in $D$ are uniformly homologous to zero in $D$, and every point of $K$ is a limit point of $D$. Then $K$ is a simple closed surface.*

Since there are two components, $S_1$ and $S_2$, in $E_3-K$, condition (1) of Theorem 3 is satisfied. Since every point of $K$ is a limit point of each of these components, condition (2) of Theorem 3 is satisfied; for, it is to be noted, condition (2) of Theorem 3 allows $P'$ to lie in that complementary domain which contains $Q$. Conditions (3) and (4) of Theorem 3 are restated in Theorem 4. Thus Theorem 4 follows at once from Theorem 3.

4. **Appendix.** The independence of the conditions of Theorem 3 is established as follows:

Condition (1): In $E_3$, using rectangular coördinates, let $K$ be the set of points $(x, y, 0)$ such that $x^2+y^2 \leqq 1$. Let $S_1$ be the set of all points for which $z>0$; let $S_2=E_3-(K+S_1)$.

Condition (2): Using spherical coördinates, let $K$ be the set of all points $(\rho, \phi, \theta)$ for which $1 \leqq \rho \leqq 2$, let $S_1$ be the set of points for which $\rho<1$, and $S_2$ the set for which $\rho>2$.

Condition (3): Using spherical coördinates, let $K'$ denote the set of points $\rho=1$. On $K'$, let $t$ denote the arc consisting of points $(1, 0, \theta)$ such that $0 \leqq \theta \leqq \pi$. On $t$, if $P$ denotes any point $(1, 0, \theta)$, let $P'$ denote the point $(1, 0, \pi-\theta)$. The set $K$ is obtained by continuously deforming $K'$ so that each point $P$ coincides with $P'$, but points not on $t$ remain distinct as before. The complement of $K$ is two domains, neither of which satisfies condition (3) of Theorem 3.

Condition (4): The torus.

The equivalence, in the case of open sets in $E_n$, of the conditions *uniformly homologous to zero* and *uniformly connected im kleinen*, is established in the following theorem:

THEOREM 5. *In order that an open set $G$ in $E_n$ should be uniformly connected im kleinen, it is necessary and sufficient that the* 0*-chains in $G$ should be uniformly homologous to zero in $G$.*

The condition is necessary. Let $\epsilon$ be any positive number. Since $G$ is uniformly connected im kleinen, there exists a positive number $\delta$ such that if $P$ and $Q$ are 0-cells of $G$ whose distance apart is less than $\delta$, there is an arc $PQ$ of $G$ whose diameter is less than $\frac{1}{4}\epsilon$.

Let $E_n - G = F$, and denote the distance between $F$ and the arc $PQ$ by $\eta_1$. Let $\eta$ be a positive number smaller than either of the numbers $\eta_1, \frac{1}{2}\epsilon$. Let $P_1(=P), P_2, \cdots, P_n(=Q)$ be points of $PQ$ such that $\rho(P_i, P_{i+1}) < \frac{1}{4}\eta (i = 1, 2, \cdots, n-1)$. The subdivision of $E_n$ may be extended so that there exist 0-cells $F_1(=P_1), F_2, \cdots, F_n(=P_n)$ such that $\rho(F_i, F_{i+1}) < \frac{1}{4}\eta$ and $\rho(F_i, P_i) < \frac{1}{4}\eta$. Let $T_i$ be a spherical neighborhood of $P_i$ of radius $\frac{1}{2}\eta$. Then clearly $F_{i+1}$ lies in $T_i$, and consequently $F_i + F_{i+1}$ bounds a 1-chain $H_i$ in $T_i$. Since no point of $F$ lies in $T_i$, $H_i \cdot F = 0$. Consequently the chain $K^1 = \sum_{i=1}^{n-1} H_i$ lies in $G$.

The diameter of $K^1$ is less than $\epsilon$. For let $x_1$ and $x_2$ be points of $K^1$. Since $x_1$ lies in some sphere $T_i$, and hence $\rho(x_1, P_i) < \frac{1}{2}\eta < \frac{1}{4}\epsilon$, and since there exists, similarly, a $P_j$ such that $\rho(x_2, P_j) < \frac{1}{4}\epsilon$, it follows at once from the fact that the diameter of $PQ$ is $< \frac{1}{4}\epsilon$, that $\rho(x_1, x_2) < \epsilon$.

The condition is sufficient. If $\epsilon$ is a positive number, there is a positive number $\delta$ such that if $x_1^0$ and $x_2^0$ are 0-cells of $G$ such that $\rho(x_1^0, x_2^0) < \delta$, then $x_1^0 + x_2^0$ bounds a chain $K^1$ of $G$ of diameter $< \frac{1}{4}\epsilon$. Let $P_1$ and $P_2$ be any two points of $G$ such that $\rho(P_1, P_2) < \frac{1}{4}\delta$. The subdivision of $E_n$ may be extended so that there exist 0-cells $y_1^0$ and $y_2^0$ such that $\rho(x_i^0, P_i) < \frac{1}{4}\delta$ $(i = 1, 2)$, and such that there are arcs $y_i P_i$ of diameter $< \frac{1}{4}\epsilon$ in $G$. Since $\rho(y_1^0, y_2^0) < \delta$, $y_1^0 + y_2^0$ bounds a chain $K^1$ of $G$ of diameter $< \frac{1}{4}\epsilon$. The set $K^1 + \sum_{i=1}^{2} y_i P_i$ contains an arc from $P_1$ to $P_2$ of diameter $< \epsilon$.

As a result of Therem 5 we can restate Theorem 4 as follows:

THEOREM 4′(Converse of the Jordan-Brouwer Separation Theorem in $E_3$). *In $E_3$, the common boundary, $K$, of two uniformly connected im kleinen domains, one of which is bounded, is a simple closed surface, provided that the Betti number* (mod 2) $R^1(E_3 - K) = 1$.

**The problem in $n$ dimensions, $n > 3$.** Regarding a converse of the Jordan-Brouwer Separation Theorem in $E_n$, where $n > 3$, the author will not hazard any guesses here, but merely indicate the likelihood that by an extension of the conditions of Theorem 4, such a converse may be obtained. Thus, condition (1) of Theorem 4 may be replaced, as a result of the Alexander duality theorem, by the condition that the Betti numbers $R^0(E_3 - K)$ and $R^i(E_3 - K)$ $(i = 1, 2, \cdots, n-2)$ shall be respectively equal to 2 and 1, and condition (2) may be replaced by the statement that the closed $i$-chains $(i = 0, 1, \cdots, n-2)$ of $D$ are uniformly homologous to zero in $D$, retaining, of course, the statement that every point of $K$ is a limit point of $D$. The validity of such a condition (2) is established by the following theorems, with which we close the present paper. (The extension of the definition of "uniformly homologous to zero" to $i$-chains where $i > 0$ should be obvious.)

THEOREM 6. *Let $M^{n-1}$ be an $(n-1)$-sphere immersed in $E_n$, and let $D$ be one of the domains of $E_n - M^{n-1}$. Then the closed $i$-chains $(n>i>0)$ of $D$ are uniformly homologous to zero in $D$.*

Suppose the conclusion of the theorem untrue. Then there is a point $P$ of $M^{n-1}$ and a contracting sequence of $i$-chains of $D$, $K_1^i, K_2^i, K_3^i, \cdots$, such that $P$ is the sequential limit point of any set of points $P_1, P_2, P_3, \cdots$, where $P_n$, for every $n$, is a point of $K_n^i$, and such that there is a positive number $\epsilon$ for which $K_n^i$, for every $n$, fails to bound any $(i+1)$-chain of $D$ of diameter less than $\epsilon$.

Let $S(P, \epsilon/2)$ be a spherical neighborhood of $P$ of diameter $\epsilon/2$. Then there is an $(n-1)$-cell, $C^{n-1}$, of $M^{n-1}$, which contains $P$ as an interior point and lies wholly in $S(P, \epsilon/2)$. Let $R$ denote a spherical neighborhood of $P$ which encloses only points of $M^{n-1}$ that lie interior to $C^{n-1}$. Denote the boundaries of $S(P, \epsilon/2)$ and $C^{n-1}$ by $F(P, \epsilon/2)$ and $C^{n-2}$, respectively. The set $C^{n-2}$ is of course an $(n-2)$-sphere immersed in $E_n$.

Let us first consider the case where $i<n-1$.

There is a number $j$, such that $K_j^i$ lies wholly in $R$. Then, denoting by $B^{n-1}$ the set of all points of $M^{n-1}$ that are not interior points of $C^{n-1}$, we have the following congruences:

$$L_1^{i+1} \equiv K_j^i \qquad (\text{mod } 2, E_n - C^{n-1}),$$
$$L_2^{i+1} \equiv K_j^i \qquad \{\text{mod } 2, R, E_n - [B^{n-1} + F(P, \epsilon/2)]\}.$$

Since $i>0$, $i+1$ is greater than 1, and since by Alexander's Theorem $X^i$ an $(n-2)$-sphere immersed in $E_n$ can be linked only by a 1-chain, the following homology holds:

$$L_1^{i+1} + L_2^{i+1} \sim 0 \qquad (\text{mod } 2, E_n - C^{n-2}).$$

As $C^{n-2}$ is the common part of $C^{n-1}$ and $B^{n-1}+F(P, \epsilon/2)$, it follows from Alexander's Corollary $W^i$ that

$$K_j^i \sim 0 \qquad [\text{mod } 2, D \cdot S(P, \epsilon/2)].$$

As this contradicts our assumption, the theorem is proved for $i<n-1$.

If $i=n-1$, the proof is trivial. Each of the chains $K_j^i$ may be assumed irreducible (i.e., $R^i(K_j^i)=2$), and thus, by Alexander's Theorem $Y$, separates $E_n$ into just two domains, one of which, $D_1$, must lie in $R$. The domain $D_1$ is an open $n$-chain bounded by $K_j^i$ and containing no point of $M^{n-1}$.

A similar proof, together with the result of Theorem 1a, yields the following general result:

THEOREM 7. *Let $M^{n-1}$ be any $(n-1)$-manifold immersed in $E_n$, $D$ a domain complementary to $M^{n-1}$, and $P$ a point of $M^{n-1}$ which can be considered as*

*interior to an $(n-1)$-cell of definition of $M^{n-1}$. Then if $U$ is a neighborhood of $P$ bounded by an $(n-1)$-sphere and enclosing only points of $M^{n-1}$ that are interior to $C^{n-1}$, there exists, in $U$, a neighborhood $V$ of $P$ such that any $i$-chain $(i=0, 1, \cdots, n-1)$ that lies in $D \cdot V$ bounds an $(i+1)$-chain of $D \cdot U$. If $M^{n-1}$ is of the type such that every one of its points can be considered as lying interior to an $(n-1)$-cell of definition of the manifold, then the $i$-chains $(i=0, 1, \cdots, n-1)$ of $D$ are uniformly homologous to zero in $D$.*

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

# ON THE ZEROS OF CERTAIN RATIONAL FUNCTIONS*

BY

MORRIS MARDEN†

## I

1. Our present note uses simple methods to obtain zero-free regions for certain functions of the type

$$F(z) = \sum \alpha_j f_j(z)$$

where $\alpha_j$ is a complex number and $f_j(z)$ is a rational function of which the approximate positions of the zeros and poles are known. As first theorem we give what is a generalization not only of some of our recent results‡ for the case that $f_j(z) = (z - a_j)^{-1}$, but also of some of Nagy's results§ for the case that the $\alpha_j$ are real and positive.

THEOREM 1. *Let $\alpha_j$ be complex numbers situated in the same given angular domain of which the vertex is the origin and the aperture is $\gamma$, $0 \leqq \gamma < \pi$. Suppose that $m$ and $n$ are two integers independent of $j$, and that, for all values of $j$ and $k$, $a_{jk}$ and $b_{jk}$ are points of the same given convex region $K$.*¶

*Then the locus of the zeros of the function*

$$F(z) = \sum_1^p \alpha_j f_j(z)$$

*where*

$$f_j(z) \equiv \frac{(z - a_{j1})(z - a_{j2}) \cdots (z - a_{jm})}{(z - b_{j1})(z - b_{j2}) \cdots (z - b_{jn})}$$

*is a region $S$ consisting of all the points at which $K$ subtends an angle $\phi$,*

$$\phi \geqq (\pi - \gamma)/(m + n).$$

Several notions entering in the statement of this theorem and its proof need to be defined. By the locus of the zeros of $F$ we mean a region outside of

---

* Presented to the Society, December 30, 1929; received by the editors March 22, 1930. The author wishes to thank Professor Pólya for the latter's many helpful suggestions and criticisms.

† National Research Fellow.

‡ M. Marden, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 363–370.

§ J. v. S. Nagy, Acta Universitatis Hungaricae Francisco-Josephinae, vol. 1 (1923), pp. 127–141.

¶ The theorem and discussion that follow obtain also for convex regions $K$ that are infinite. Any of the points $M_1$, $M_2$, $N_1$, $N_2$ entering in the definition of "subtended angle" as given below may then be at infinity.

which no point is a zero of $F$ and inside of which any point can be made a zero of $F$ through a suitable choice of the numbers $\alpha_j$ and of the points $a_{jk}$ and $b_{jk}$. By the symbol $\sphericalangle P_1PP_2$ we shall mean throughout this paper the angle measured in the counterclockwise sense from the ray $PP_1$ to the ray $PP_2$. Finally, by the angle subtended by the region $K$ at a point $P$ outside of or on the boundary of $K$, we mean the angle $\phi$, $\phi \leqq \pi$, such that for every pair of points $M_1$, $M_2$ in $K$ (with the subscripts so chosen that $\sphericalangle M_1PM_2$ does not exceed $\pi$)

$$\sphericalangle M_1PM_2 \leqq \phi,$$

and such that for at least two points $N_1$, $N_2$ in $K$

$$\sphericalangle N_1PN_2 = \phi.$$

If $P$ is an interior point of $K$, the angle subtended by $K$ at $P$ will be taken as $\pi$.

On the basis of the preceding definitions let us investigate some of the properties of the region $S$, the ensemble of all the points at which $K$ subtends an angle $\phi$, $\phi \geqq \delta$, where $\delta \equiv (\pi-\gamma)/(m+n)$. Clearly $S$ contains $K$. If $P$ is an arbitrary point of $S$, and $Q$ an arbitrary point of $K$, every point of the line-segment $PQ$ is also included in the region $S$. This fact is evident if $P$ lies in $K$, for the region $K$ was assumed to be convex. If $P$ is not in $K$ and $R$ is an arbitrary point of the segment $PQ$, we shall denote by $M_1'$ and $M_2'$ the points of intersection with the line $N_1N_2$ of the lines through $R$ parallel to the lines $PN_1$ and $PN_2$. As $K$ is a convex region, the segment $N_1N_2$ and therefore the points $M_1'$ and $M_2'$ lie in $K$. This means, if $\phi$ and $\psi$ are the angles subtended by $K$ at $P$ and $R$ respectively, that

$$\psi \geqq \phi \geqq \delta, \tag{1}$$

and that hence $R$ is a point of $S$. In other words, *the region $S$ is star-shaped with respect to every point of $K$.**

Inequality (1) leads us to a further descriptive property of $S$. The region $S$, which for $\gamma = m = n-1 = 0$ is $K$ itself, expands as $\gamma$ and $m+n$ take on larger values and becomes the entire plane when either $\gamma$ approaches $\pi$ or $m+n$ becomes infinite.

2. To prove Theorem 1, we must establish the two propositions (a) that no point outside of $S$ is a zero of $F$ and (b) that any point in $S$ is a zero of $F$ for appropriately chosen numbers $\alpha_j$ and functions $f_j(z)$.

(a) Let $P:z$ be an arbitrary point outside of $K$; let $\phi$ be the angle subtended at $P$ by $K$; let $N_1$ and $N_2$ be two points of $K$ such that $\sphericalangle N_1PN_2 = \phi$, and let $\pi+\tau$ be the angle made by the ray $PN_1$ with the positive real axis.

* A proof of this property may also be found on p. 128 of Nagy's article.

Since the points $a_{jk}$ and $b_{jk}$ lie in the region $K$, the quantities $(z-a_{jk})$ and $(z-b_{jk})^{-1}$ are vectors drawn to $P$ having the properties

$$\tau \leqq \arg (z - a_{jk}) \leqq \tau + \phi,$$
$$-(\tau + \phi) \leqq \arg (z - b_{jk})^{-1} \leqq -\tau.$$

Hence,

$$(m - n)\tau - n\phi \leqq \arg f_j(z) \leqq (m - n)\tau + m\phi.$$

If, consequently, we suppose that for all $j$

$$\kappa \leqq \arg \alpha_j \leqq \kappa + \gamma,$$

we find that

$$(m - n)\tau - n\phi + \kappa \leqq \arg \alpha_j f_j(z) \leqq (m - n)\tau + m\phi + \kappa + \gamma.$$

For the point $P$ to be a zero of $F$ it is necessary that the difference between the extreme members of the latter inequality be at least $\pi$. That is to say, the point $P$ cannot be a zero of $F$ unless $\phi \geqq \delta$, $\delta \equiv (\pi - \gamma)/(m+n)$, i.e., unless $P$ lies in $S$.

(b) For the second part of the proof, let us assume $P:\zeta$ to be an arbitrary point of $S$ and $\phi$ to be the angle subtended at $P$ by $K$. Because $\phi \geqq \delta$, there exist in $K$ two points $Q_1:c_1$, $Q_2:c_2$ for which $\sphericalangle Q_1PQ_2 = \delta$. Let $d_1$ and $d_2$ denote the distances from $P$ of the points $Q_1$ and $Q_2$ respectively, and let $\sigma$ denote the angle formed by the ray $PQ_1$ with the positive real axis. Let furthermore

$$t_1 \equiv [(\zeta - c_1)d_1^{-1}]^{m+n} \quad \text{and} \quad t_2 \equiv [(\zeta - c_2)d_2^{-1}]^{m+n}e^{\gamma i}.$$

Then, as

$$\arg t_1 = (m + n)\sigma,$$
$$\arg t_2 = (m + n)(\sigma + \delta) + \gamma = \pi + (m + n)\sigma,$$

the vectors $t_1$ and $t_2$ are equal and opposite and thus

$$t_1 + t_2 = 0.$$

Now consider the function

$$G(z) = \frac{d_2^n (z - c_1)^m}{d_1^m (z - c_2)^n} + \frac{e^{\gamma i} d_1^n (z - c_2)^m}{d_2^m (z - c_1)^n}.$$

It has the point $P$ as a zero and it satisfies all the hypotheses of the theorem. Proposition (b) is therefore affirmed and with it the proof of Theorem 1 completed.

3. We shall now present a number of special cases and corollaries to the theorem.

(a) On our setting $\gamma = 0$, Theorem 1 reduces to Nagy's general theorem

(Nagy, p. 131), and on our placing $\gamma=m=n-1=0$ it becomes Gauss's theorem for the derivative of a polynomial.

(b) If $f_j(z)=z-a_j$, the zero of $F$ may be thought of as the "center of gravity" of the system of complex "masses," the mass $\alpha_j$ being at the point $a_j$. Thus, if all the points $a_j$ lie in the convex region $K$, their center of gravity lies in $S$, the region consisting of all points at which $K$ subtends an angle $\phi$, $\phi \geqq \pi-\gamma$. In particular, if all the $\alpha_j$ are real, $\gamma=0$ and $S$ coincides with $K$.

(c) Let us specialize $K$ to be in turn the ray $\theta=0$ (the positive real axis), the line-segment $AB$, and a circle $C$ of radius $r$.

In the first instance, the region $S$ is the angular domain $|\theta| \leqq \pi-\delta$.

In the second instance, $S$ is a region bounded by two circular arcs. These arcs are symmetric in the line $AB$ and cut one another in the points $A$ and $B$ at an angle of $2\delta$. The points $A$ and $B$ are however not to be included in $S$. If $\gamma-\pi/2=m=n-1=0$, the region $S$ becomes the interior and circumference of the circle having segment $AB$ as diameter—a result found also on page 366 of our previous note.

In the third instance, $S$ is the interior and circumference of the circle concentric with $C$ and of radius $r/\sin(\delta/2)$. For $\gamma-\pi/2=m=n-1=0$, this result coincides with that on page 366 of our previous note.

For the special case that $K$ is a polygonal region we refer to the discussion on pages 128–9 of Nagy's article.

(d) By virtue of Theorem 1, we may modify Nagy's generalization of Jentzsch's theorem so that it will read as follows:

*Let $\alpha_1$ and $\alpha_2$ be two complex numbers situated in the same given angular domain of which the vertex is the origin and the aperture is $\gamma$, $0 \leqq \gamma < \pi$. Let $f(z)$ be a polynomial of degree $n$ and $g_1(z)$ and $g_2(z)$ polynomials each of degree at most $m$, $m \leqq n-1$. Then, if all the zeros of the two functions*

$$F_1(z) = f(z) - g_1(z) \text{ and } F_2(z) = f(z) - g_2(z)$$

*lie in the same given convex region $K$, the zeros of the function*

$$F(z) = f(z) - \frac{\alpha_1 g_1(z) + \alpha_2 g_2(z)}{\alpha_1 + \alpha_2}$$

*will lie in a region $S$ consisting of all points at which $K$ subtends an angle $\phi$, $\phi \geqq (\pi-\gamma)/(m+1)$.*

We have given this theorem without proof, as a proof can be constructed in close analogy with that given by Nagy (his article, p. 138) for his generalization of Jentzsch's theorem.

(e) Let us next suppose that

$$f_j(z) \equiv \left(1 - \frac{\lambda_j}{n} z\right)^{-n}$$

where the $\lambda_j$ are real numbers with $0 < \lambda_j < \lambda_{j+1}$. As we saw in part (a) of this section, there are no zeros in the angular domain which is bounded by the rays $L_1'$ and $L_2'$ drawn from the point $z = n/\lambda_p$ at an inclination of $\delta$ with the positive real axis, and which does not include the points $z = n/\lambda_j$. The lines $L_1'$ and $L_2'$ have $y$-intercepts of magnitude $l' = n\lambda_p^{-1} \tan \delta$. Now, when $n$ becomes infinite, the function $f_j(z)$ approaches $e^{\lambda_j z}$; the quantity $l'$ approaches $l = \lambda_p^{-1}(\pi - \gamma)$ and the lines $L_1'$ and $L_2'$ become the parallels $L_1$ and $L_2$ to the $x$-axis at a distance $l$ above and below the $x$-axis. In the strip between $L_1$ and $L_2$ there are consequently no zeros of the function

$$\Phi = \sum_1^p \alpha_j e^{\lambda_j z}.$$

This result may, of course, be deduced with ease directly from the function $\Phi$.

(f) Finally, we notice the result described in Theorem 1 to be independent of the number of terms entering in the sum $F$ and therefore to be valid for certain infinite sums. This fact permits us, for instance, to state the following theorem.

*Given the regular curve $\Gamma$ and the convex region $K$. Let $\alpha(t)$ be a complex function such that for all points $t$ on $\Gamma$*

$$\kappa \leqq \arg [\alpha(t)dt] \leqq \kappa + \gamma < \kappa + \pi.$$

*Let $a_j(t)$ and $b_j(t)$ be complex functions such that for all $j$ and for all points $t$ on $\Gamma$ the points $u_j$ and $v_j$, $u_j \equiv a_j(t)$ and $v_j \equiv b_j(t)$, lie in $K$. Then the integral*

$$I(z) = \int_\Gamma \alpha(t) f(z, t) dt,$$

*where*

$$f(z, t) \equiv \frac{\prod_{j=1}^{m} (z - a_j(t))}{\prod_{j=1}^{n} (z - b_j(t))},$$

*if it exists, has no zeros in the region $S$ consisting of all the points at which $K$ subtends an angle $\phi$, $\phi < (\pi - \gamma)/(m+n)$.*

This theorem may be derived directly from Theorem 1 by dividing $\Gamma$ into $p$ parts by means of the successive points $t_j$, by setting $\Delta t_j \equiv t_{j+1} - t_j$, $\alpha_j \equiv \alpha(t_j)\Delta t_j$, and $f_j(z) \equiv f(z, t_j)$, and by finally letting $p$ become infinite.

The theorem has as immediate consequence one concerning the $n$th derivative of certain analytic functions. If $K$ is bounded by the curve $\Gamma$ and the function $\psi(z)$ is analytic outside of $K$ (including the point at infinity) and continuous on $\Gamma$, the $n$th derivative of $\psi(z)$ has at a point $z$ outside of $K$ the value

$$\psi^{(n)}(z) = -\frac{n!}{2\pi i}\int_\Gamma \frac{\psi(t)dt}{(t-z)^{n+1}}.$$

As the latter integral is of the type considered above, the ensuing theorem is true.

*Given a convex region $K$, bounded by a closed, regular curve $\Gamma$. Let $\psi(z)$ be a function which is analytic outside of $K$ (including the point at infinity) and continuous on $\Gamma$, and which at every point $t$ of $\Gamma$ has the property*

$$\kappa \leqq \text{arc } [\psi(t)dt] \leqq \kappa + \gamma < \kappa + \pi.$$

*Then the nth derivative of $\psi(z)$ has no zeros in the region $T$ consisting of all the points at which $K$ subtends an angle $\phi$, $\phi < (\pi-\gamma)/(n+1)$.*

The zero-free regions $S$ and $T$ of the two preceding theorems do not, however, give the largest such regions. This is clear from the fact that the best approximation in Theorem 1 is attained only for sums $F$ of just two terms. (Cf. §3 (a).) When, for instance, $n=0$ and $K$ is a circle, a larger zero-free region than $T$ may be obtained directly. In this case

$$\psi(t)dt \equiv it\psi(t)d\theta$$

where $\theta \equiv \text{arc } t$. As $\psi(t)\cdot t$ is thus limited to an angular domain of opening less then $\pi$, it cannot increase by even $2\pi$ when the point $t$ makes a complete circuit around $\Gamma$. Thus the entire exterior of $K$ is a zero-free region of $\psi^{(n)}(z)$.

## II

4. Suppose $F_1$ and $F_2$ are two functions of the type considered in Part I of our present note. What can be said of the zeros of the difference $\Phi = F_1 - F_2$? This question will be answered only in the case that $m+n=1$. We make this restriction expressly for the purpose of obtaining results which will have the same order of simplicity as those of Part I.

The theorem given below on the zeros of $\Phi$ involves the notion of the "inner angle" at a point $P$ of two non-overlapping convex regions $K$ and $L$. To explain this term we shall suppose $P$ to be an arbitrary point of the plane; $\theta$ and $\phi$ to be the angles subtended at $P$ by the regions $K$ and $L$ respectively, and $M_1$, $M_2$ to be a pair of points in $K$ and $N_1$, $N_2$ a pair of points in $L$ for which

$$\sphericalangle M_1PM_2 = \theta \quad \text{and} \quad \sphericalangle N_1PN_2 = \phi.$$

Suppose the angles $M_1PM_2$ and $N_1PN_2$ do not overlap one another. Then the smaller of the two angles $M_2PN_1$ and $N_2PM_1$ is what we mean by the "inner angle at $P$ of the regions $K$ and $L$." The notion of inner angle will not be defined in the case that the angles $M_1PM_2$ and $N_1PN_2$ overlap one another.

Our theorem concerning the zeros of $\Phi$ is then the following:

THEOREM 2. *Let $\alpha_j$ and $\beta_k$ be complex numbers situated in the same given angular domain of which the vertex is the origin and the aperture is $\gamma$, $0 \leqq \gamma < \pi$. Let $K$ and $L$ be two non-overlapping convex regions, in the first of which lie the points $a_j$, $j=1, 2, \cdots, p$, and in the second of which lie the points $b_k$, $k=1, 2, \cdots, q$. Suppose that $T$ is the ensemble of all points at which the inner angle of $K$ and $L$ is $\psi$, $\psi > \gamma$.*

*Then the locus of the zeros of the function**

$$F(z) = \sum_{j=1}^{p} \frac{\alpha_j}{z - a_j} - \sum_{k=1}^{q} \frac{\beta_k}{z - b_k}$$

*is the region complementary† to the region $T$.*

5. The proof of this theorem, like that of Theorem 1, falls into two distinct parts.

(a) If we set

$$G \equiv \sum_{j=1}^{p} \frac{\alpha_j}{a_j - z} \quad \text{and} \quad H \equiv \sum_{k=1}^{q} \frac{\beta_k}{b_k - z},$$

we have as necessary condition for the point $P:z$ to be a zero of $F$ that the angular domain in which the vector $G$ may range overlap the one in which the vector $H$ may range.

We observe that multiplication of $F$ by the constant factor $e^{ki}$ does not affect the zeros of $F$. Without loss of generality we may therefore assume that

$$-\pi/2 < -\gamma/2 \leqq \arc \alpha_j \leqq \gamma/2 < \pi/2$$

and

$$-\pi/2 < -\gamma/2 \leqq \arc \beta_k \leqq \gamma/2 < \pi/2.$$

Let us draw the lines $PQ_1$, $PQ_2$, $PR_1$, $PR_2$; $PM_1'$, $PM_2'$, $PN_1'$, $PN_2'$,

* The theorem is also true for the function

$$f(z) = \sum_{j=1}^{p} \alpha_j(z - a_j) - \sum_{k=1}^{q} \beta_k(z - b_k)$$

whose zero is the "center of gravity" of the system of complex "masses," the mass $\alpha_j$ being at the point $a_j$ and the mass $-\beta_k$ being at the point $b_k$. For $\alpha_j$ and $\beta_j$ real and positive, the function $F(z)$ is the logarithmic derivative of a rational function.

† We are using this adjective to mean the region composed of all the points of the plane which are not in $T$.

$PQ_1'$, $PQ_2'$, $PR_1'$, $PR_2'$, where the notation of §4 is used, where $\sphericalangle Q_1PM_1 = \sphericalangle M_2PQ_2 = \sphericalangle R_1PN_1 = \sphericalangle N_2PR_2 = \gamma/2$, and where the primed letters denote points obtained by reflecting the corresponding unprimed lettered points in the line through $P$ parallel to the real axis.

As the points $a_j$ lie in $K$ and the points $b_k$ in $L$, the quantities $(a_j - z)^{-1}$ and $(b_k - z)^{-1}$ are vectors drawn from $P$, the first in the angle $M_2'PM_1'$ and the second in the angle $N_2'PN_1'$. The vectors $\alpha_j(a_j - z)^{-1}$ and $\beta_k(b_k - z)^{-1}$ are respectively in the angles $Q_2'PQ_1'$ and $R_2'PR_1'$. The vector $G$ therefore ranges in the angle $Q_2'PQ_1'$; the vector $H$, in the angle $R_2'PR_1'$.

Now, the angles $Q_2'PQ_1'$ and $R_2'PR_1'$ will overlap one another when and only when the angles $Q_1PQ_2$ and $R_1PR_2$ overlap one another. If the angles $M_1PM_2$ and $N_1PN_2$ overlap one another, the same will be true of the angles $Q_1PQ_2$ and $R_1PR_2$. If the angles $M_1PM_2$ and $N_1PN_2$ do not overlap one another, the angles $Q_1PQ_2$ and $R_1PR_2$ will overlap one another in and only in the case that the inner angle at $P$ does not exceed $\gamma$. This means that the zeros of $F$ must always lie in the region complementary to the region $T$.

(b) We need yet to show that conversely any point $P:\zeta$ of the complement to $T$ can be made a zero of $F$ through a suitable choice of the numbers $\alpha_j$ and $\beta_k$ and of the points $a_j$ and $b_k$.

As $P$ is a point of the complement of $T$, the angles $Q_1PQ_2$ and $R_1PR_2$ necessarily overlap one another. We shall suppose that they so overlap that the line $PR_1$ lies in the angular domain $Q_1PQ_2$. Let $PS$, $PU$ and $PV$ be rays drawn in the angular domains $R_1PQ_2$, $M_1PM_2$ and $N_1PN_2$ respectively, these rays being so chosen that $\sphericalangle UPS = \sphericalangle SPV = \gamma/2$. Now on the line $PU$ exists a point $a$ of the region $K$ and on the line $PV$ exists a point $b$ of the region $L$. Denote the distance from $P$ of the point $a$ by $c$ and that of the point $b$ by $d$. Then the function

$$F = \frac{ce^{-\gamma i/2}}{z - a} - \frac{de^{\gamma i/2}}{z - b}$$

has $P$ as a zero, for at $P$ it is a sum of two equal and opposite vectors. This function being one which satisfies the hypotheses of the theorem, the proof of Theorem 2 is now complete.

6. In this concluding section we shall try to throw light on the nature of the region $T$ of Theorem 2 through a study of some special cases.

Suppose, first, that $K$ consists merely of the point $A$, and $L$ merely of the point $B$. The region $T$, as the locus of the points $P$ where the angle $APB$ exceeds $\gamma$, is the interior of the region $\Delta$ bounded by the two circular arcs which, symmetric in the line $AB$, cut one another in $A$ and $B$ at an angle of $2\gamma$. The same region obtains for the case that $K$ and $L$ are two segments or

half-rays along the same line, and $A$ and $B$ are the points of $K$ and $L$ respectively which are nearest together.

Suppose, secondly, that $K$ is a circle with center at the point $C_1$: $(-a, 0)$ and of radius $r_1$, and $L$ is a circle with center at the point $C_2$:$(a, 0)$ and of radius $r_2$. Suppose the line-segment $C_1C_2$ to intersect circle $K$ in a point $A$ and circle $L$ in a point $B$.

To secure an upper limit to the region $T$, we concentrate at $A$ all the points $a_j$ and at $B$ all the points $b_k$. The new locus of the zeros of $F$ is clearly contained in the old and accordingly the new region $T'$ encloses the old $T$. The region $T'$ is precisely the interior of the region $\Delta$ described in the second paragraph of this section (§6).

For an accurate determination of the region $T$ we need to concern ourselves, on account of symmetry, only with the upper half-plane.

Let us first suppose $\gamma$ to be small. Then in the upper half-plane exist just one point $Q$ on the circle $K$ and just one point $R$ on the circle $L$ at which the inner angle is $\gamma$. The locus of the points outside of $K$ and $L$ at which the inner angle of $K$ and $L$ is $\gamma$ is then an arc $\sigma$ of a curve which lies in the region $\Delta$ and connects the points $Q$ and $R$. If $T_1$ denotes the interior of the closed curve formed by joining together the arc $\sigma$ with the arc $QA$ of circle $K$, with the arc $RB$ of circle $L$ and with the line-segment $AB$, the region $T_1$ will comprise all the points of the upper half-plane at which the inner angle of $K$ and $L$ is defined and is greater than $\gamma$. The region $T$ is therefore composed of $T_1$ and of the reflection of $T_1$ in the line $AB$.

To obtain the equation of the curve $S$ of which the arc $\sigma$ forms a part, we shall assume $P:(x, y)$ to be a point outside of $K$ and $L$, and $M$ and $N$ to be the points of contact with $K$ and $L$ respectively of those tangents from $P$ for which angle $MPN=\gamma$. In terms of the parameter $\theta \equiv \sphericalangle AC_1M$, the equations of $S$ are

$$(a + x)\cos\theta + y\sin\theta = r_1,$$
$$(a - x)\cos(\gamma + \theta) - y\sin(\gamma + \theta) = r_2.$$

The parameter eliminated, the curve $S$ is represented by the single equation

$$[(x^2 + y^2 - a^2)\sin\gamma - 2ay]^2 - [r_1(a - x)\sin\gamma + y(r_1\cos\gamma + r_2)]^2$$
$$- [x(r_2 + r_1\cos\gamma) + r_1y\sin\gamma + a(r_2 - r_1\cos\gamma)]^2 = 0.$$

Hence $S$ is a bi-circular quartic. As application of the usual method shows, it is a bi-circular quartic with its two singular foci coincident in the point $z=ia\csc\gamma$.*

* In fact this curve has been considered in a slightly different connection by Walsh, who shows that it is a cartesian oval. See Quarterly Journal of Mathematics, vol. 50 (1924), pp. 154–165.

What occurs when $\gamma$ approaches zero? The equation of $S$ becomes

$$[4a^2 - (r_1 + r_2)^2]y^2 = [x(r_1 + r_2) + a(r_2 - r_1)]^2,$$

which factors into

$$\pm\, y(4a^2 - (r_1 + r_2)^2)^{1/2} = x(r_1 + r_2) + a(r_2 - r_1).$$

The bi-circular quartic thus degenerates into the two inner common tangents of the circles $K$ and $L$, and the points $Q$ and $R$ become the points of contact (in the upper half-plane) of these tangents with $K$ and $L$ respectively. The region $T$ is now the interior of the infinite region bounded by the line-segment $AB$, the arc $QA$ of circle $K$, the arc $RB$ of circle $L$ and the rays along these inner common tangents from the points $Q$ and $R$ indefinitely outwards.

This result agrees with the deduction which, by allowing $m_1$ and $m_2$ to vary arbitrarily, one can make from a theorem due to Walsh.* We shall word this theorem as follows:

Let the numbers $\alpha_j$ and $\beta_k$ be real and positive, and the sums

$$m_1 = \sum_1^p \alpha_j \quad \text{and} \quad m_2 = \sum_1^q \beta_k$$

be such that $m_1 > m_2$. Suppose the points $a_j$ to lie in a circle $K$ with center $\xi_1$ and of radius $r_1$, and the points $b_k$ to lie in a circle $L$ with center $\xi_2$ and of radius $r_2$. Then all the finite zeros of the function

$$\sum_1^p \frac{\alpha_j}{z - a_j} - \sum_1^q \frac{\beta_k}{z - b_k}$$

lie in a circle whose center is $(m_1 - m_2)^{-1}\ (m_1\xi_2 - m_2\xi_1)$ and whose radius is $(m_1 - m_2)^{-1}(m_1 r_2 + m_2 r_1)$.

So much for the case that $\gamma$ is small. Let us now examine that in which $\gamma$ has nearly the value $\pi$. In this case there exists no point on the circle $K$, the circle $L$, or the segment $AB$ at which the inner angle equals $\gamma$. The curve $S$ is a closed curve lying in the upper half-plane within the region $\Delta$. Its equation has still the form given above. As to the region $T$, it consists of the interior of $S$ and of the interior of the reflection of $S$ in the line $AB$.

When $\gamma$ approaches $\pi$, the region $\Delta$ collapses into coincidence with the line-segment $AB$. As at no point of the segment $AB$ is the inner angle of $K$ and $L$ equal to $\gamma$ (we are assuming the product $r_1 r_2 \neq 0$), the closed curve $S$ must shrink to a point on the approach of $\gamma$ to a certain value $\gamma_1$, $\gamma_1 < \pi$. For this value $\gamma = \gamma_1$, the region $T$, as the interior of $S$ and of the reflection of $S$ in $AB$, must disappear, and the locus of the zeros of $F$, as the region

* J. L. Walsh, these Transactions, vol. 22 (1921), bottom of p. 114.

complementary to $T$, must become the entire plane. The entire plane is therefore the locus of the zeros of $F$ for values of $\gamma$ between $\gamma_1$ and $\pi$.

It is now intuitively clear what deformations the region $T$ undergoes as $\gamma$ varies from zero to $\pi$. The end points $Q$ and $R$ of the arc $\sigma$ move down the circles $K$ and $L$ towards the points $A$ and $B$ respectively. Then $Q$ and $R$ draw closer together along the segment $AB$ until they reach coincidence. They coincide at the point of the segment $AB$ where the inner angle of $K$ and $L$ takes on its maximum $\gamma_0$. As $\gamma$ then changes from $\gamma_0$ to $\gamma_1$, the arc $\sigma$ remains a closed curve in the upper half of the region $\Delta$. When finally $\gamma$ becomes equal to $\gamma_1$, the oval shrinks to a point, and then and thereafter $T$ contains no points.

EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE,
ZURICH, SWITZERLAND

# THE BEHAVIOR OF BOUNDS AND OSCILLATIONS OF SEQUENCES OF FUNCTIONS UNDER REGULAR TRANSFORMATIONS*

BY

RALPH PALMER AGNEW

**Introduction.** Much has been written during the past thirty years on summability† of sequences of constants. The subject of summability of sequences of functions is largely unexplored.‡

Very recently Gillespie and Hurwitz§ discovered that any bounded sequence of real continuous functions, defined over a closed compact set $A$, which converges to a continuous limit function is uniformly summable by a regular transformation totally equivalent to a regular transformation with a triangular matrix. This fact creates an interest in summability of sequences of functions. In 1925 Hurwitz¶ discussed the behavior of what we shall call ultimate bounds of sequences of real constants under real regular transformations, and more recently he‖ has considered the behavior of oscillations of complex and of real sequences of constants under complex and real regular transformations with triangular matrices.

In this paper, regular transformations with triangular matrices, which have been conspicuous in the theory of summability, will be applied to sequences of functions. We will consider the behavior of ultimate bounds of real sequences under real transformations, and of continuous oscillation and convergence and uniform oscillation and convergence of complex and real sequences under complex and real transformations; ultimate bounds, oscillations, and convergence being considered (1) over a set as a whole, (2) at a single point of a set, and (3) at all points and limit points of a set.**

---

* Presented to the Society, April 18, 1930; received by the editors April 2, 1930.

† For a bibliography of the subject, see Smail, *History and Synopsis of the Theory of Summable Infinite Processes*, University of Oregon Press, 1925.

‡ However, it is well known that the Cesàro and certain other transformations preserve uniform convergence of bounded sequences of functions.

§ D. C. Gillespie and W. A. Hurwitz, these Transactions, vol. 32, pp. 527–543.

¶ W. A. Hurwitz, Proceedings of the London Mathematical Society, (2), vol. 26, pp. 231–48.

‖ W. A. Hurwitz, American Journal of Mathematics.

** In a paper to appear in the American Journal of Mathematics, entitled *The behavior of mean square oscillation and convergence under regular transformations*, the writer further extends the application of regular transformations to sequences of functions.

We recall that a transformation with a triangular matrix is a sequence-to-sequence transformation of the form

$$\sigma_n = a_{n1}s_1 + a_{n2}s_2 + \cdots + a_{nn}s_n$$

where the $a_{nk}$ are constants, and that a transformation is said to be *regular* when it carries every convergent sequence $\{s_n\}$ into a sequence $\{\sigma_n\}$ which converges to the same value. Such a transformation is said to be real when $a_{nk}$ is real for all $n$ and $k$; otherwise it is complex. The following ten conditions to which we shall refer repeatedly in this paper are catalogued together for convenience:

$C_1$ : $\sum_{k=1}^{n} |a_{nk}|$ is bounded for all $n$;

$C_2$ : for each $k$, $\lim_{n\to\infty} a_{nk} = 0$;

$C_3$ : $\lim_{n\to\infty} \sum_{k=1}^{n} a_{nk} = 1$;

$C_4$ : $\lim_{n\to\infty} \sum_{k=1}^{n} |a_{nk}| = 1$;

$C_5$ : for each $k$, $a_{nk} = 0$ for almost* all $n$:

$C_6$ : $\sum_{k=1}^{n} a_{nk} = 1$ for almost all $n$;

$C_7$ : $a_{nk} \geqq 0$ for almost all $k$;

$C_8$ : for each $k$, $a_{nk} \geqq 0$ for almost all $n$;

$C_9$ : for each $k$, $a_{nk} \leqq 0$ for almost all $n$;

$C_{10}$: $a_{nk} \geqq 0$ for almost all $n$.

The reader will find it helpful to think of each of the ten conditions in terms of the triangular matrix of which the $n$th row is $a_{n1}, a_{n2}, \cdots, a_{nn}$, and the $k$th column is $a_{kk}, a_{k+1,k}, a_{k+2,k}, \cdots$. Thus $C_6$ indicates that there is a certain row such that the sum of the elements in any lower row is 1, $C_7$ indicates that all elements to the right of a certain fixed column are non-negative, $C_{10}$ indicates that all elements below a certain fixed row are non-negative, etc.

Let the symbol $(T)$ represent a regular transformation with a triangular matrix. By the Silverman-Toeplitz theorem, in order that a complex (or real) transformation may be regular when applied to complex (or real) sequences, $C_1$, $C_2$, and $C_3$ are necessary and sufficient. Hence $(T)$, complex

* I.e., $a_{nk}=0$ except for at most a finite number, depending on $k$, of values of $n$.

or real, satisfies $C_1$, $C_2$, and $C_3$. The conditions $C_4$ and $C_7$ were used by Hurwitz,* and $C_5$, $C_6$, $C_8$, $C_9$, and $C_{10}$ are used for the first time in this paper.

The functions $f(x)$ and $s(x)$ considered in this paper are regarded as being defined over a non-empty set $A$ of any abstract metric space; however any reader not interested in this degree of generality may regard the set $A$ as being in euclidean space† of one or more dimensions. In all sections where complex functions are considered, an interesting special case is obtained by regarding $A$ as being a set in a euclidean plane so that a point $x$ of $A$ determines two coördinates, say $\xi$ and $\eta$, and a complex function $f(x)$ may be written $u(\xi, \eta)+iv(\xi, \eta)$; if one cares to demand that $u$ and $v$ shall satisfy the Cauchy-Riemann differential equations, then $f(x)$ is an analytic function of the complex variables $(\xi+i\eta)$. Throughout this paper, $f(x)$ is finite for each value of $x$ for which it is defined, but the symbols $+\infty$ and $-\infty$ are admitted as limits and bounds.

The wordings of the theorems of the various sections of this paper have been made as similar as possible to make essential differences apparent and to facilitate the proofs. The reader may note that alternative statements of some of the theorems are desirable; for example, Theorem 7.12 is equivalent to the following:

*Any regular transformation with a triangular matrix carries every sequence which is bounded and uniformly convergent over a given set into a sequence which converges uniformly over the set to the same limit function.*

## CHAPTER I. BOUNDS OF REAL SEQUENCES

In this chapter, all sequences and transformations considered are real.

Two facts, established by W. A. Hurwitz,‡ to which we shall have occasion to refer in this chapter are stated in the following lemmas:

LEMMA 1.01. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} \sigma_n \leqq \limsup_{n\to\infty} s_n$$

*for all sequences* $\{s_n\}$ *of constants for which the right member is finite,* $C_7$ *is necessary and sufficient.*

LEMMA 1.02. *If* $(T)$ *fails to satisfy* $C_4$, *then there is a bounded sequence* $\{s_n\}$ *of constants such that*

$$\limsup_{n\to\infty} \sigma_n > \limsup_{n\to\infty} s_n.$$

* Proceedings of the London Mathematical Society, loc. cit.

† In any euclidean space, any bounded set is compact and conversely.

‡ Proceedings of the London Mathematical Society, vol. 26 (1925), pp. 232–237.

## 1. Ultimate bounds of sequences over a set

Let a sequence $\{f_n(x)\}$ be defined over a set $A$. Then $G(f_n; A)$ is the least upper bound of the element $f_n(x)$ of the sequence over the set $A$; and $\limsup_{n\to\infty} G(f_n; A)$ may appropriately be called the *ultimate least upper bound of the sequence* $\{f_n(x)\}$ *over the set* $A$. Analogously $g(f_n; A)$ is the greatest lower bound of $f_n(x)$ over $A$, and $\liminf_{n\to\infty} g(f_n; A)$ is the *ultimate greatest lower bound of* $\{f_n(x)\}$ *over* $A$.

The seven theorems 1.110–1.123 give necessary and sufficient conditions that $(T)$ shall not increase ultimate least upper bounds of sequences over a set.

Theorem 1.110. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A) \tag{1.1101}$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

The condition $C_4$ is necessary, for it follows from Lemma 1.02 that it is necessary for (1.1101) when bounded sequences of constants are considered. The condition $C_4$ is sufficient. Given an admissible sequence $\{s_n(x)\}$, choose $M$ so that $|s_n(x)| < M$ over $A$ for all $n$ and let $q$ be any number greater than $\limsup G(s_n; A)$. Choose an index $p$ such that $G(s_n; A) < q$ for $n > p$; then $s_n(x) < q$ over $A$ for $n > p$. Let $b_{nk} = \frac{1}{2}(|a_{nk}| + a_{nk}) \geqq 0$ and $c_{nk} = \frac{1}{2}(|a_{nk}| - a_{nk}) \geqq 0$ so that $|a_{nk}| = b_{nk} + c_{nk}$ and $a_{nk} = b_{nk} - c_{nk}$. Then for $n > p$ and each $x$ in $A$

$$\sigma_n(x) = \sum_{k=1}^{p} a_{nk}s_k - \sum_{k=p+1}^{n} c_{nk}s_k + \sum_{k=p+1}^{n} b_{nk}s_k, \tag{1.1102}$$

$$\sigma_n(x) \leqq M\sum_{k=1}^{p} |a_{nk}| + M\sum_{k=p+1}^{n} c_{nk} + q\sum_{k=p+1}^{n} b_{nk}. \tag{1.1103}$$

Since the right member of (1.1103) is independent of $x$, we have for $n > p$

$$G(\sigma_n; A) \leqq M\sum_{k=1}^{p} |a_{nk}| + M\sum_{k=p+1}^{n} c_{nk} + q\sum_{k=p+1}^{n} b_{nk}. \tag{1.1104}$$

Using $C_4$ and the regularity of $(T)$ we obtain the following five relations the first three of which hold for each fixed $k$:

$$\lim_{n\to\infty} |a_{nk}| = 0, \quad \lim_{n\to\infty} c_{nk} = 0, \quad \lim_{n\to\infty} b_{nk} = 0, \quad \lim_{n\to\infty} \sum_{k=1}^{n} c_{nk} = 0, \quad \lim_{n\to\infty} \sum_{k=1}^{n} b_{nk} = 1;$$

and due to these relations it follows from (1.1104) that $\limsup G(\sigma_n; A) \leqq q$.

Since $q$ is any number greater than lim sup $G(s_n; A)$, (1.1101) follows and $C_4$ is sufficient.

Two theorems, the proofs of which may be made to depend on those of other theorems, are the following:

THEOREM 1.111. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded above over* $A$ *for all* $n$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

THEOREM 1.112. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded below over* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *is necessary and sufficient.*

Since $A$ is a finite set, the condition $s_n(x)$ *is bounded above over* $A$ *for all* $n$ is automatically satisfied by every sequence $\{s_n(x)\}$ for which lim sup $G(s_n; A)$ is finite. Therefore Theorem 1.111 is equivalent to Theorem 1.113 and may be proved simultaneously with it; and Theorem 1.112 follows from Theorem 1.110.

THEOREM 1.113. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

Necessity follows from Lemma 1.01. The condition is sufficient. Let $q$ be any number greater than lim sup $G(s_n; A)$, and choose an index $p$ such that $G(s_n; A) < q$ for $n > p$ and also $a_{nk} \geqq 0$ for $k > p$. Since $A$ is finite, $s_n(x)$ is bounded over $A$ for each $n$; hence there is a constant $M$ such that $|s_n(x)| < M$ over $A$ for $n = 1, 2, 3, \cdots, p$. Therefore for $n > p$ and each $x$ in $A$

$$\sigma_n(x) = \sum_{k=1}^{p} a_{nk}s_k + \sum_{k=p+1}^{n} a_{nk}s_k \leqq M \sum_{k=1}^{p} |a_{nk}| + q \sum_{k=p+1}^{n} a_{nk},$$

and

$$G(\sigma_n; A) \leqq M \sum_{k=1}^{p} |a_{nk}| + q \sum_{k=p+1}^{n} a_{nk}.$$

Since $(T)$ is regular, lim sup $G(\sigma_n; A) \leqq q$, and sufficiency follows.

THEOREM 1.121. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A) \tag{1.1211}$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above over* $A$ *for all* $n$, *such that the right member is finite,* $C_{10}$ *is necessary and sufficient.*

We shall show that $C_{10}$ is necessary by showing that $C_7$ and $C_8$, which taken together are equivalent to $C_{10}$, are both necessary. The necessity of $C_7$ follows from Lemma 1.01. To show that $C_8$ is necessary, we shall suppose that $(T)$ fails to satisfy $C_8$ and construct a sequence $\{s_n(x)\}$ satisfying the conditions of the theorem for which (1.1211) is denied. From a denial of $C_8$ it follows that there is a value of $k$, say $\lambda$, and a sequence $\{n_\alpha\}$ of indices such that $\lim n_\alpha = +\infty$ and $a_{n_\alpha\lambda} < 0$ for $\alpha = 1, 2, 3, \cdots$. Let $\{x_\alpha\}$ be a sequence of distinct points of $A$, and define the sequence $\{s_n(x)\}$ over $A$ as follows: $s_n(x) \equiv 0$, $n \neq \lambda$; $s_\lambda(x) = 0$, $x \neq x_1, x_2, x_3, \cdots$; and $s_\lambda(x_\alpha) = 1/a_{n_\alpha\lambda}$, $\alpha = 1, 2, 3, \cdots$. Note that $s_n(x)$ is bounded above (by 0) over $A$ for all $n$ and that $\limsup G(s_n; A) = \liminf g(s_n; A) = 0$.* However for $n_\alpha \geqq \lambda$, $\sigma_{n_\alpha}(x_\alpha) = 1$, $\alpha = 1, 2, 3, \cdots$, so that $G(\sigma_{n_\alpha}; A) \geqq 1$, $\alpha = 1, 2, 3, \cdots$, and $\limsup G(\sigma_n; A) \geqq 1$. Thus, for this sequence $\{s_n(x)\}$, (1.1211) is denied and the necessity follows. To establish sufficiency of $C_{10}$, choose $M$ such that $s_n(x) < M$ over $A$ for all $n$ and define $q$ and $p$ as in the preceding theorems. Let $p$ be increased if necessary so that $a_{nk} \geqq 0$ for $n > p$. Then for $n > p$ and each $x$ in $A$

$$\sigma_n(x) = \sum_{k=1}^{p} a_{nk}s_k + \sum_{k=p+1}^{n} a_{nk}s_k \leqq M \sum_{k=1}^{p} a_{nk} + q \sum_{k=p+1}^{n} a_{nk};$$

and since $(T)$ is regular, $\limsup G(\sigma_n; A) \leqq q$ and sufficiency follows.

THEOREM 1.122. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded below over* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *and,* $C_9$ *are necessary and sufficient.*

Necessity of $C_4$ follows from Theorem 1.110. The necessity of $C_9$ is established here, where $s_n(x)$ is not assumed bounded above, by using the method used in Theorem 1.121, where $s_n(x)$ was not assumed bounded below,

* The fact that $\liminf g(s_n; A)$ isfinite will be used in the proof of Theorem 1.321.

to prove that $C_8$ was a necessary condition. To establish sufficiency, choose $\mu$ such that $s_n(x) > \mu$ over $A$ for all $n$ and define $q$ and $p$ as usual. Due to $C_9$ there is an index $N \geqq p$ such that $a_{nk} \leqq 0$ for $k = 1, 2, 3, \cdots, p$ when $n > N$. Hence for $n > N$ and $x$ in $A$

$$\sigma_n(x) \leqq \mu \sum_{k=1}^{p} a_{nk} - \mu \sum_{k=p+1}^{n} c_{nk} + q \sum_{k=p+1}^{n} b_{nk}$$

where $c_{nk}$ and $b_{nk}$ are the non-negative constants defined in Theorem 1.110. Using $C_4$ and the regularity of $(T)$, we find that lim sup $G(\sigma_n; A) \leqq q$ and sufficiency follows.

THEOREM 1.123. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that the right member is finite,* $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

That $C_{10}$ and $C_9$ are necessary follows from Theorems 1.121 and 1.122. But $C_{10}$ and $C_9$ are equivalent to $C_{10}$ and $C_5$; thus necessity is proved. To establish sufficiency, define $q$ and $p$ as usual and increase $p$ if necessary so that $a_{nk} \geqq 0$ for $n > p$. Due to $C_5$ there is an index $N \geqq p$ such that $a_{nk} = 0$, $k = 1, 2, 3, \cdots, p$, for $n > N$. Then we have for $n > N$ and each $x$ in $A$

$$\sigma_n(x) = \sum_{k=1}^{p} a_{nk}s_k + \sum_{k=p+1}^{n} a_{nk}s_k = \sum_{k=p+1}^{n} a_{nk}s_k \leqq q \sum_{k=p+1}^{n} a_{nk};$$

hence lim sup $G(\sigma_n; A) \leqq q$ and sufficiency follows.

The seven theorems 1.210–1.223 give necessary and sufficient conditions that $(T)$ shall not decrease ultimate greatest lower bounds of sequences over a set. Rewriting Theorems 1.110, 1.112, 1.111, 1.113, 1.122, 1.121, and 1.123 with $s_n(x)$ replaced by $-s_n(x)$ and hence $\sigma_n(x)$ replaced by $-\sigma_n(x)$, and taking into account facts such as lim sup $G(-f_n; A) = -$lim inf $g(f_n; A)$ and if $-s_n(x)$ is bounded above over $A$ for all $n$ then $s_n(x)$ is bounded below over $A$ for all $n$, we obtain respectively the seven theorems:

THEOREM 1.210. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n; A) \geqq \liminf_{n\to\infty} g(s_n; A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 1.211. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded above over* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *is necessary and sufficient.*

THEOREM 1.212. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded below over* $A$ *for all* $n$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

THEOREM 1.213. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

THEOREM 1.221. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above over* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *and* $C_9$ *are necessary and sufficient.*

THEOREM 1.222. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded below over* $A$ *for all* $n$, *such that the right member is finite,* $C_{10}$ *is necessary and sufficient.*

THEOREM 1.223. *In order that* $(T)$ *may be such that*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that the right member is finite,* $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

The seven theorems 1.310–1.323 give necessary and sufficient conditions that a transformation shall *neither* increase ultimate least upper bounds *nor* decrease ultimate greatest lower bounds of sequences.

We may combine Theorems 1.110 and 1.210 to obtain

THEOREM 1.310. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n;\ A) \leqq \limsup_{n\to\infty} G(s_n;\ A)$$

*and*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

The three following theorems may be proved simultaneously.

THEOREMS 1.311, 1.312. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n;\ A)] \leqq \limsup_{n\to\infty} G(s_n;\ A)$$

*and*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, *such that the right members are finite,* $C_4$ *is necessary and sufficient.*

THEOREM 1.313. *In order that* $(T)$ *may be such that*

$$\limsup_{n\to\infty} G(\sigma_n;\ A) \leqq \limsup_{n\to\infty} G(s_n;\ A)$$

*and*

$$\liminf_{n\to\infty} g(\sigma_n;\ A) \geqq \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, *such that the right members are finite,* $C_4$ *is necessary and sufficient.*

Since $A$ is a finite set, the sequences $\{s_n(x)\}$ considered in each of the three theorems are precisely the sequences $\{s_n(x)\}$ which are bounded over $A$ for all $n$; hence each of the three theorems follows from Theorem 1.310.

THEOREM 1.321. *In order that* $(T)$ *may be such that*

(1.3211) $$\limsup_{n\to\infty} G(\sigma_n;\ A) \leqq \limsup_{n\to\infty} G(s_n;\ A)$$

*and*

(1.3212) $$\liminf_{n\to\infty} g(\sigma_n;\ A) \liminf_{n\to\infty} g(s_n;\ A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above over* $A$ *for all* $n$, *such that the right members are finite,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

The necessity of $C_4$ is contained in Theorem 1.310. That $C_8$ is necessary for (1.3211) under the present conditions was established in the course of the proof of Theorem 1.121. The proof of Theorem 1.122 is easily made to show that $C_9$ is necessary for (1.1221) even when lim inf $g(s_n; A)$ is supposed finite, and we find on replacing $s_n(x)$ by $-s_n(x)$ that $C_9$ is necessary for (1.3212) under the present conditions. Since $C_8$ and $C_9$ are necessary, $C_5$ is necessary. To establish sufficiency, let $Q$ be any number greater than lim sup $G(s_n; A)$ and $q$ be any number less than lim inf $g(s_n; A)$. Choose an index $p$ such that $G(s_n; A)<Q$ and $g(s_n; A)>q$ for $n>p$. Using $C_5$, choose an index $N \geqq p$ such that $a_{nk}=0$, $k=1, 2, 3, \cdots, p$, for $n>N$. Then, using the notation of Theorem 1.110, we may write for $n>N$ and each $x$ in $A$

$$-Q\sum_{k=p+1}^{n} c_{nk} + q\sum_{k=p+1}^{n} b_{nk} \leqq \sigma_n(x)$$

$$\leqq -q\sum_{k=p+1}^{n} c_{nk} + Q\sum_{k=p+1}^{n} b_{nk}.$$

Using $C_4$ and the regularity of $(T)$ we find from the preceding relation that $q \leqq \liminf g(\sigma_n; A)$ and $\limsup G(\sigma_n; A) \leqq Q$; these relations imply (1.3211) and (1.3212) and sufficiency is proved.

Rewriting the preceding theorem with $s_n(x)$ replaced by $-s_n(x)$, we obtain

THEOREM 1.322. *In order that $(T)$ may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*and*

$$\liminf_{n\to\infty} g(\sigma_n; A) \geqq \liminf_{n\to\infty} g(s_n; A)$$

*for every sequence $\{s_n(x)\}$, defined over an infinite set $A$ and bounded below over $A$ for all $n$, such that the right members are finite, $C_4$ and $C_5$ are necessary and sufficient.*

THEOREM 1.323. *In order that $(T)$ may be such that*

$$\limsup_{n\to\infty} G(\sigma_n; A) \leqq \limsup_{n\to\infty} G(s_n; A)$$

*and*

$$\liminf_{n\to\infty} g(\sigma_n; A) \geqq \liminf_{n\to\infty} g(s_n; A)$$

*for every sequence $\{s_n(x)\}$, defined over an infinite set $A$, such that the right members are finite, $C_4$ and $C_5$ are necessary and sufficient.*

Necessity is contained in Theorem 1.321 and sufficiency may be established verbatim as in Theorem 1.321.

## 2. Ultimate bounds of sequences at a point

Let a sequence $\{f_n(x)\}$ be defined over a set $A$ and let $x_0$ be a point of $A^0$ where $A^0$ is the set consisting of the points and limit points of $A$. For each sequence $\{x_i\}$ of points of $A$ with the limit $x_0$, form

$$\limsup_{n\to\infty,\, i\to\infty} f_n(x_i) = v \quad \text{and} \quad \liminf_{n\to\infty,\, i\to\infty} f_n(x_i) = w;$$

the least upper bound of all such $v$ is denoted by $\Gamma(x_0; \{f_n\}, A)$* and may be called the *ultimate least upper bound of the sequence* $\{f_n(x)\}$ *at the point* $x_0$; the greatest lower bound of all such $w$ is denoted by $\gamma(x_0; \{f_n\}, A)$ and may be called the *ultimate greatest lower bound of the sequence* $\{f_n(x)\}$ *at the point* $x_0$.

The seven theorems 2.110–2.123 give necessary and sufficient conditions that $(T)$ shall not increase ultimate least upper bounds of sequences at a point.

Theorem 2.110. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A) \tag{2.1101}$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

Necessity is established as in Theorem 1.110. To prove sufficiency, choose $M$ so that $|s_n(x)| < M$ over $D$ for all $n$, let $q$ be any number greater than $\Gamma(x_0; \{s_n\}, A)$, and let $\{x_i\}$ be any sequence of points of $A$ with the limit $x_0$. Choose an index $p$ such that $s_n(x_i) < q$ for $i > p$, $n > p$ and such that $x_i$ is in $D$ for $i > p$. Then we may write, for $x = x_i$, $n > p$, $i > p$, the relations (1.1102) and (1.1103). Using $C_4$ and the regularity of $(T)$ we find that

$$\limsup_{n\to\infty,\, i\to\infty} \sigma_n(x_i) \leqq q;$$

hence $\Gamma(x_0; \{\sigma_n\}, A) \leqq q$, (2.1101) follows, and the theorem is proved.

Theorem 2.111. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded above over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

* Hahn, *Theorie der reellen Funktionen*, p. 232.

THEOREM 2.112. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded below over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *is necessary and sufficient.*

Since $x_0$ is an isolated point of $A$, the condition $s_n(x)$ *is bounded above over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$ is automatically satisfied by every sequence $\{s_n(x)\}$ for which $\Gamma(x_0; \{s_n\}, A)$ is finite. Therefore Theorem 2.111 is equivalent to Theorem 2.113 and may be proved simultaneously with it; and Theorem 2.112 follows from Theorem 2.110.

THEOREM 2.113. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A) \tag{2.1131}$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point, such that the right member is finite,* $C_7$ *is necessary and sufficient.*

If $\{x_i\}$ and $\{x_i'\}$ are any two sequences of points of $A$ with the limit $x_0$, then there is an index $p$ such that $x_i$ and $x_i'$ are $x_0$ for $i > p$. Hence

$$\limsup_{n\to\infty,\, i\to\infty} \sigma_n(x_i) = \limsup_{n\to\infty} \sigma_n(x_0),$$

and

$$\limsup_{n\to\infty,\, i\to\infty} s_n(x_i') = \limsup_{n\to\infty} s_n(x_0)$$

and (2.1131) is equivalent to the condition

$$\limsup_{n\to\infty} \sigma_n(x_0) \leqq \limsup_{n\to\infty} s_n(x_0).$$

But $\{s_n(x_0)\}$ is a sequence of constants; thus the necessity and sufficiency of $C_7$ follows from Lemma 1.01.

THEOREM 2.121. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_{10}$ *is necessary and sufficient.*

The necessity proof is an adaptation of that of Theorem 1.121, the essential change being that the sequence $\{x_\alpha\}$ is for the present purpose taken to be a sequence of distinct points of $A$ with the limit $x_0$. The suffi-

ciency proof is a modification of that of Theorem 1.121 precisely as the sufficiency proof of Theorem 2.110 is a modification of that of Theorem 1.110

Similar proofs furnish the two following theorems:

THEOREM 2.122. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded below over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *and* $C_9$ *are necessary and sufficient.*

THEOREM 2.123. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that the right member is finite,* $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

The seven theorems 2.210–2.223 give necessary and sufficient conditions that $(T)$ shall not decrease ultimate greatest lower bounds of sequences at a point. Rewriting Theorems 2.110, 2.112, 2.111, 2.113, 2.122, 2.121, and 2.123 with $s_n(x)$ replaced by $-s_n(x)$, and taking into account facts such as $\Gamma(x_0; \{-f_n\}, A) = -\gamma(x_0; \{f_n\}, A)$ we obtain respectively the seven theorems:

THEOREM 2.210. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 2.211. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded above over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *is necessary and sufficient.*

THEOREM 2.212. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded below over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_7$ *is necessary and sufficient.*

THEOREM 2.213. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point, such that the right member is finite,* $C_7$ *is necessary and sufficient.*

THEOREM 2.221. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_4$ *and* $C_9$ *are necessary and sufficient.*

THEOREM 2.222. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded below over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right member is finite,* $C_{10}$ *is necessary and sufficient.*

THEOREM 2.223. *In order that* $(T)$ *may be such that*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that the right member is finite,* $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

The seven theorems 2.310–2.323 give necessary and sufficient conditions that $(T)$ shall *neither* increase ultimate least upper bounds *nor* decrease ultimate greatest lower bounds of sequences at a point. The proofs of these theorems are modifications of the proofs of the corresponding theorems of §1.

THEOREM 2.310. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*and*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREMS 2.311, 2.312. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*and*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right members are finite,* $C_4$ *is necessary and sufficient.*

THEOREM 2.313. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*and*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point, such that the right members are finite,* $C_4$ *is necessary and sufficient.*

THEOREMS 2.321, 2.322. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*and*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that the right members are finite,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 2.323. *In order that* $(T)$ *may be such that*

$$\Gamma(x_0; \{\sigma_n\}, A) \leqq \Gamma(x_0; \{s_n\}, A)$$

*and*

$$\gamma(x_0; \{\sigma_n\}, A) \geqq \gamma(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that the right members are finite,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

### 3. ULTIMATE BOUNDS AT ALL POINTS AND LIMIT POINTS OF A SET

In comparing the theorems of this section with those of §1, the following facts, given as lemmas, should be noted.

LEMMA 3.01. *If*

$$\limsup_{n\to\infty} G(f_n; A) \leqq \limsup_{n\to\infty} G(F_n; A)$$

*where $A$ contains more than one point, then it is not necessarily true that $\Gamma(x; \{f_n\}, A) \leqq \Gamma(x; \{F_n\}, A)$ over $A$.*

This lemma is established by considering the following simple sequences: $f_n(x_1)=1$, and $f_n(x)=0$ for $x\neq x_1$; $F_n(x_2)=1$, and $F_n(x)=0$ for $x\neq x_2$; where $x_1$ and $x_2$ are two points of $A$. The first statement holds but the second fails for $x=x_1$. A change of the signs of the elements of the sequences establishes a corresponding lemma for ultimate greatest lower bounds.

LEMMA 3.02. *If $\Gamma(x; \{f_n\}, A) \leqq \Gamma(x; \{F_n\}, A)$ over $A^0$ and $A$ is compact,* then*

$$\limsup_{n\to\infty} G(f_n; A) \leqq \limsup_{n\to\infty} G(F_n; A).$$

Suppose $A$ is compact and lim sup $G(f_n; A) >$ lim sup $G(F_n; A)$; we shall show that the hypothesis is contradicted. Choose a number $\theta$ such that lim sup $G(f_n; A) > \theta >$ lim sup $G(F_n; A)$. Then there is a sequence $\{n_\alpha\}$ such that $G(f_{n_\alpha}, A) > \theta$ for $\alpha = 1, 2, 3, \cdots$, and hence a sequence $\{x_\alpha\}$ of points of $A$ such that $f_{n_\alpha}(x_\alpha) > \theta$. Since $A$ is compact, let $\{x_\alpha'\}$ be a subsequence of $\{x_\alpha\}$ with a limit, say $x_0$; then $\Gamma(x_0; \{f_n\}, A) \geqq \theta$. It is easy to show that lim sup $G(F_n; A) \geqq \Gamma(x; \{F_n\}, A)$ over $A^0$. Combining inequalities, we have $\Gamma(x_0; \{f_n\}, A) > \Gamma(x; \{F_n\}, A)$ over $A^0$, the hypothesis is denied, and the lemma is proved. A corresponding lemma for greatest lower bounds follows.

The seven theorems 3.110–3.123 give necessary and sufficient conditions that $(T)$ shall not increase ultimate least upper bounds of sequences at the points and limit points of a set.

From Theorems 2.110, 2.111, 2.112 and 2.113, we obtain respectively the following four theorems.

THEOREM 3.110. *In order that $(T)$ may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence $\{s_n(x)\}$, defined over an arbitrary set $A$ and bounded over $A$ for all $n$, $C_4$ is necessary and sufficient.*

---

* The condition "$A$ is compact" cannot be removed. To show this, let $\{x_n\}$ be a sequence of points of $A$ without limit points and define: $f_n(x_n)=2$, $f_n(x)=0$ for $x\neq x_n$; and $F_n(x)=1$ over $A$. Then $\Gamma(x; \{f_n\}, A)=0$ over $A^0$, $\Gamma(x; \{F_n\}, A)=1$ over $A^0$, lim sup $G(f_n; A)=2$, and lim sup $G(F_n; A)=1$ and the memla fails.

THEOREM 3.111. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded above over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A$,* $C_7$ *is necessary and sufficient.*

THEOREM 3.112. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded below over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A$, $C_4$ *is necessary and sufficient.*

THEOREM 3.113. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points, such that the right member is finite for each* $x$ *in* $A$, $C_7$ *is necessary and sufficient.*

Noting that any set is made up of limit points and isolated points, we obtain the necessity of the condition of the following theorem from Theorem 2.121 and the sufficiency from Theorems 2.121 and 2.111.

THEOREM 3.121. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A^0$, $C_{10}$ *is necessary and sufficient.*

Similarly we obtain from Theorems 2.122 and 2.112, and from Theorems 2.123 and 2.113, the two following theorems:

THEOREM 3.122. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded below over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A^0$, $C_4$ *and* $C_9$ *are necessary and sufficient.*

* But not necessarily bounded over $A$.

THEOREM 3.123. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that the right member is finite for each* $x$ *in* $A^0$, $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

The seven theorems 3.210–3.223, obtained from the preceding by replacing $s_n(x)$ by $-s_n(x)$, give necessary and sufficient conditions that $(T)$ shall not decrease ultimate greatest lower bounds of sequences at the points and limit points of a set.

THEOREM 3.210. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 3.211. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded above over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A$, $C_4$ *is necessary and sufficient.*

THEOREM 3.212. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded below over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A$, $C_7$ *is necessary and sufficient.*

THEOREM 3.213. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points, such that the right member is finite for each* $x$ *in* $A$, $C_7$ *is necessary and sufficient.*

THEOREM 3.221. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A^0$, $C_4$ *and* $C_9$ *are necessary and sufficient.*

THEOREM 3.222. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded below over* $A$ *for all* $n$, *such that the right member is finite for each* $x$ *in* $A^0$, $C_{10}$ *is necessary and sufficient.*

THEOREM 3.223. *In order that* $(T)$ *may be such that*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that the right member is finite for each* $x$ *in* $A^0$, $C_{10}$ *and* $C_5$ *are necessary and sufficient.*

The seven theorems 3.310–3.323, obtained from the last seven theorems of §2, give necessary and sufficient conditions that $(T)$ shall *neither* increase ultimate least upper bounds *nor* decrease ultimate greatest lower bounds of sequences at the points and limit points of a set.

THEOREM 3.310. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*and*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREMS 3.311, 3.312. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A$$

*and*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded above (below) over* $A$ *for all* $n$, *such that the right members are finite for each* $x$ *in* $A$, $C_4$ *is necessary and sufficient.*

THEOREM 3.313. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A$$

*and*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points, such that the right members are finite for each* $x$ *in* $A$, $C_4$ *is necessary and sufficient.*

THEOREMS 3.321, 3.322. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*and*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above (below) over* $A$ *for all* $n$, *such that the right members are finite for each* $x$ *in* $A^0$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 3.323. *In order that* $(T)$ *may be such that*

$$\Gamma(x; \{\sigma_n\}, A) \leqq \Gamma(x; \{s_n\}, A) \text{ over } A^0$$

*and*

$$\gamma(x; \{\sigma_n\}, A) \geqq \gamma(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that the right members are finite for each* $x$ *in* $A^0$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

## CHAPTER II. CONTINUOUS OSCILLATON AND CONVERGENCE

In this chapter the sequences and transformations considered may be, except in cases where a specific statement to the contrary is made, either real or complex.

We include in the following lemma a fact taken from a paper by W. A. Hurwitz.*

LEMMA 4.01. *If* $(T)$ *fails to satisfy* $C_4$, *then there is a bounded sequence* $\{s_n\}$ *of constants such that*

$$\limsup_{m\to\infty,\, n\to\infty} |\sigma_m - \sigma_n| > \limsup_{m\to\infty,\, n\to\infty} |s_m - s_n| ;$$

*if* $(T)$ *is real,* $\{s_n\}$ *may be taken real.*

The four following lemmas may be proved together.

LEMMAS 4.02, 4.03. *Let* $A$ *be an infinite set and let* $\{x_\alpha\}$ *be a sequence of distinct points in* $A$. *In case* $A$ *has a limit point* $x_0$, *let the sequence* $\{x_\alpha\}$ *be a sequence with the limit* $x_0$. *If* $(T)$, *real or complex, fails to satisfy* $C_5$, *then there is a real sequence* $\{s_n(x)\}$, *bounded above (below) over* $A$, *such that*

$$\limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |\sigma_m(x_i) - \sigma_n(x_j)| > \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |s_m(x_i') - s_n(x_j')| = 0$$

*where* $\{x_\alpha'\}$ *is any sequence of points of* $A$.

* American Journal of Mathematics.

LEMMAS 4.04, 4.05. *Under the hypotheses of Lemmas* 4.02, 4.03 *there is a real sequence* $\{s_n(x)\}$, *bounded above (below) over* $A$, *such that*

$$\limsup_{m\to\infty,\, n\to\infty,\, i\to\infty} |\sigma_m(x_i) - \sigma_n(x_i)| > \limsup_{m\to\infty,\, n\to\infty,\, i\to\infty} |s_m(x_i') - s_n(x_i')| = 0$$

*where* $\{x_\alpha'\}$ *is any sequence of points of* $A$.

From a denial of $C_5$ it follows that there is a value of $k$, say $\lambda$, and a sequence $\{n_\alpha\}$ of indices such that $\lim n_\alpha = +\infty$ and $a_{n_\alpha,\lambda} \neq 0$ for $\alpha = 1, 2, 3, \cdots$. Define the sequence $\{s_n(x)\}$ as follows: $s_n(x) = 0$ over $A$ for $n \neq \lambda$; $s_\lambda(x) = 0$ for $x \neq x_1, x_2, x_3, \cdots$; and $s_\lambda(x_\alpha) = (-1)^h / |a_{n_\alpha,\lambda}|$ where $h = 1$ $(h = 2)$. Evidently the sequence $\{s_n(x)\}$ is bounded above or below over $A$ according as $h$ is 1 or 2; and since $s_n(x) = 0$ over $A$ for $n > \lambda$

$$\limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |s_m(x_i') - s_n(x_j')| = \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty}} |s_m(x_i') - s_n(x_i')| = 0$$

where $\{x_\alpha'\}$ is any sequence of points of $A$. But for $n > \lambda$, $\sigma_n(x_\alpha) = a_{n\lambda} s_\lambda(x_\alpha)$ and we find for $n_\alpha > \lambda$, $n_\beta > \lambda$ that

$$|\sigma_{n_\alpha}(x_\alpha)| = 1 \text{ and } |\sigma_{n_\beta}(x_\alpha)| = |a_{n_\beta,\lambda}| / |a_{n_\alpha,\lambda}|;$$

hence

$$\limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |\sigma_m(x_i) - \sigma_n(x_j)| \geqq \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty}} |\sigma_m(x_i) - \sigma_n(x_i)|$$

$$\geqq \limsup_{\alpha\to\infty,\, \beta\to\infty} \big| |\sigma_{n_\alpha}(x_\alpha)| - |\sigma_{n_\beta}(x_\alpha)| \big| \geqq \limsup_{\alpha\to\infty,\, \beta\to\infty} |1 - |a_{n_\beta,\lambda}| / |a_{n_\alpha,\lambda}|| = +\infty$$

and the lemmas are proved.

## 4. OSCILLATION OVER A SET

Let $\{f_n(x)\}$ be defined over any set $A$. Then the *oscillation of the sequence* $\{f_n(x)\}$ *over the set* $A$ is denoted by $\Omega(\{f_n\}, A)$ and is defined as follows: For each sequence $\{x_i\}$ of points of $A$, form

$$\limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |f_m(x_i) - f_n(x_j)| = v;$$

the least upper bound of all such $v$ is $\Omega(\{f_n\}, A)$.

An interesting connection between the oscillation of a real sequence over a set and the ultimate bounds of the sequence over the set is given in the following.

LEMMA 4.06. *If* $\{f_n(x)\}$ *is a real sequence defined over any set* $A$, *then*

$$\Omega(\{f_n\}, A) = \limsup_{n\to\infty} G(f_n; A) - \liminf_{n\to\infty} g(f_n; A) \tag{4.061}$$

*whenever the right member has a meaning; otherwise** $\Omega(\{f_n\}, A) = +\infty$.

To prove this lemma, it suffices to consider three cases: first, $f_n(x)$ is bounded over $A$ for all sufficiently great $n$; second, lim sup $G(f_n; A) = +\infty$; third, lim inf $g(f_n; A) = -\infty$. In the first case, there is a sequence $\{x_i\}$ of points of $A$ such that $f_m(x_{2m}) \leqq G(f_m, A) < f_m(x_{2m}) + 1/m$ and $f_n(x_{2n+1}) \geqq g(f_n; A) > f_n(x_{2n+1}) - 1/n$ for all sufficiently great $m$ and $n$; for this sequence,

$$\limsup_{m\to\infty,\, n\to\infty} |f_m(x_{2m}) - f_n(x_{2n+1})| = \limsup_{m\to\infty} G(f_m; A) - \liminf_{n\to\infty} g(f_n; A),$$

so that $\Omega(\{f_n\}, A) \geqq \limsup G(f_n; A) - \liminf g(f_n; A)$. But for any sequence $\{x_i\}$ of points of $A$, $f_m(x_i) \leqq G(f_m; A)$ and $f_n(x_j) \geqq g(f_n; A)$ so that $\Omega(\{f_n\}, A) \leqq \limsup G(f_n; A) - \liminf g(f_n; A)$. Combining inequalities, we have (4.061) for the first case. In case lim sup $G(f_n; A) = +\infty$, let $x_0$ be a point of $A$, choose an index $n_1$ such that $G(f_{n_1}; A) > f_1(x_0) + 1$, and choose a point $x_1$ such that $f_{n_1}(x_1) > f_1(x_0) + 1$; and in general choose $n_p$ such that $G(f_{n_p}; A) > f_p(x_{p-1}) + p$ and choose $x_p$ such that $f_{n_p}(x_p) > f_p(x_{p-1}) + p$. Then

$$\lim_{p\to\infty} |f_{n_p}(x_p) - f_p(x_{p-1})| = +\infty, \quad \Omega(\{f_n\}, A) = +\infty,$$

and the lemma is proved for the second case. In case lim inf $g(f_n; A) = -\infty$, we can choose sequences $\{n_p\}$ and $\{x_p\}$ such that $f_{n_p}(x_p) < f_p(x_{p-1}) - p$, thus obtaining $\Omega(\{f_n\}, A) = +\infty$ and completing the proof of the lemma.

It is of interest to note specifically that this lemma enables us to use the latter theorems of §1 to furnish easy proofs of the sufficiency (but not the necessity) of the conditions of Theorems 4.112, 4.122, 4.123, 4.132, and 4.133.

The eight theorems 4.111–4.133 give necessary and sufficient conditions that $(T)$ shall not increase oscillations of sequences over a set.

THEOREM 4.111. *In order that* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A) \tag{4.1111}$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

Necessity follows from Lemma 4.01. To establish sufficiency, choose $M$ such that $|s_n(x)| < M$ over $A$ for all $n$, $B$ such that $\sum_{k=1}^{n} |a_{nk}| < B$ for

* The right member fails to have a meaning when and only when the terms are both $+\infty$ or both $-\infty$.

all $n$, and let $\sum_{k=1}^{n}|a_{nk}|=B_n$. Let $\{x_i\}$ be any sequence of points of $A$, and let $q$ be any number greater than $\Omega(\{s_n\}, A)$; then there is an index $p$ such that

$$(4.1112)\quad |s_\mu(x_i)-s_\nu(x_j)|<q \text{ for } \mu\geqq p,\ \nu\geqq p,\ i\geqq p,\ j\geqq p.$$

We may write, for $m>p$, $n>p$, $i>p$, $j>p$, the identity

$$\sigma_m(x_i)-\sigma_n(x_j)=\sum_{k=1}^{m}a_{mk}s_k(x_i)-\sum_{k=1}^{n}a_{nk}s_k(x_j)$$

$$=\sum_{k=1}^{p}a_{mk}s_k(x_i)-\sum_{k=1}^{p}a_{nk}s_k(x_j)+\sum_{\mu=p+1}^{m}a_{m\mu}s_\mu(x_i)-\sum_{\nu=p+1}^{n}a_{n\nu}s_\nu(x_j),$$

and, on noting that the last two terms of the last member are identically equal to

$$\Big(\sum_{\mu=p+1}^{m}a_{m\mu}s_\mu(x_i)\Big)\Big(1-\sum_{\nu=p+1}^{n}a_{n\nu}\Big)-\Big(\sum_{\nu=p+1}^{n}a_{n\nu}s_\nu(x_j)\Big)\Big(1-\sum_{\mu=p+1}^{m}a_{m\mu}\Big)$$

$$+\sum_{\mu=p+1}^{m}\sum_{\nu=p+1}^{n}a_{m\mu}a_{n\nu}[s_\mu(x_i)-s_\nu(x_j)],$$

obtain the identity

$$(4.1113)\quad \begin{aligned}\sigma_m(x_i)-\sigma_n(x_j)=&\sum_{k=1}^{p}a_{mk}s_k(x_i)-\sum_{k=1}^{p}a_{nk}s_k(x_j)\\&+\Big(\sum_{\mu=p+1}^{m}a_{m\mu}s_\mu(x_i)\Big)\Big(1-\sum_{\nu=p+1}^{n}a_{n\nu}\Big)\\&-\Big(\sum_{\nu=p+1}^{n}a_{n\nu}s_\nu(x_j)\Big)\Big(1-\sum_{\mu=p+1}^{m}a_{m\mu}\Big)\\&+\sum_{\mu=p+1}^{m}\sum_{\nu=p+1}^{n}a_{m\mu}a_{n\nu}[s_\mu(x_i)-s_\nu(x_j)].\end{aligned}$$

The absolute values of the first four terms of the right member are respectively less than or equal to

$$M\sum_{k=1}^{p}|a_{mk}|,\quad M\sum_{k=1}^{p}|a_{nk}|,\quad MB\Big|1-\sum_{\nu=p+1}^{n}a_{n\nu}\Big| \text{ and } MB\Big|1-\sum_{\mu=p+1}^{m}a_{m\mu}\Big|,$$

each of which, due to the regularity of $(T)$, approaches 0 as $m$ and $n$ become infinite; hence

$$(4.1114)\qquad \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |\sigma_m(x_i) - \sigma_n(x_j)| = \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} \left| \sum_{\mu=p+1}^{m} \sum_{\nu=p+1}^{n} a_{m\mu}a_{n\nu}[s_\mu(x_i) - s_\nu(x_j)] \right|.$$

From (4.1112) we obtain for $m>p$, $n>p$, $i>p$, $j>p$,

$$(4.1115)\qquad \left| \sum_{\mu=p+1}^{m} \sum_{\nu=p+1}^{n} a_{m\mu}a_{n\nu}[s_\mu(x_i) - s_\nu(x_j)] \right| \leqq q \sum_{\mu=p+1}^{m} \sum_{\nu=p+1}^{n} |a_{m\mu}a_{n\nu}| \leqq qB_mB_n;$$

hence

$$(4.1116)\qquad \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |\sigma_m(x_i) - \sigma_n(x_j)| \leqq \limsup_{m\to\infty,\, n\to\infty} (qB_mB_n)$$

and using $C_4$ we have

$$(4.1117)\qquad \limsup_{\substack{m\to\infty,\, n\to\infty \\ i\to\infty,\, j\to\infty}} |\sigma_m(x_i) - \sigma_n(x_j)| \leqq q.$$

Since $\{x_i\}$ is any sequence of points of $A$, $\Omega(\{\sigma_n\}, A) \leqq q$; this leads to (4.1111) and the theorem is proved.

THEOREM 4.112. *In order that a real* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

This theorem is established by noting that the proof of the preceding theorem is undisturbed by supposing $(T)$ and the $\{s_n(x)\}$ sequences to be real.

THEOREM 4.121. *In order that* $(T)$ *may be such that*

$$(4.1211)\qquad \Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, $C_4$ *is necessary and sufficient.*

Necessity follows from Theorem 4.111. The condition is sufficient. If $\Omega(\{s_n\}, A) = +\infty$, no proof is required, for (4.1211) is automatically satisfied. If $\Omega(\{s_n\}, A)$ is finite, then, since $A$ is a finite set, $s_n(x)$ is bounded over $A$ for

all $n$ and sufficiency follows from Theorem 4.111. This argument, based on Theorem 4.112, establishes

THEOREM 4.122. *In order that a real* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, $C_4$ *is necessary and sufficient.*

From Theorems 4.112 and 4.122 we obtain

THEOREM 4.123. *In order that a real* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a finite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 4.131. *In order that* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

Necessity of $C_4$ follows from Lemma 4.01, and that of $C_5$ from Lemma 4.02 (or 4.03), since for the $\{s_n(x)\}$ sequences there defined, $\Omega(\{s_n\}, A)=0$ while $\Omega(\{\sigma_n\}, A)>0$. No proof of sufficiency is required if $\Omega(\{s_n\}, A)=+\infty$. If $\Omega(\{s_n\}, A)$ is finite, let $q$ be any greater number, let $\{x_i\}$ be any sequence of points of $A$, and choose an index $p$ such that (4.1112) is satisfied. Then $|s_n(x_i)-s_p(x_p)|<q$ for $n\geqq p, i\geqq p$; hence there is a constant $M$ such that $|s_n(x_i)|<M$ for $n\geqq p$, $i\geqq p$. Using $C_5$, choose an index $N\geqq p$ such that $a_{nk}=0$, $k=1, 2, 3, \cdots, p$ for $n>N$. Then, referring to the identity (4.1113), we see that, for $m>N$, $n>N$, $i>p$, $j>p$, the first two terms of the right member vanish and the third and fourth terms approach 0 as $m$ and $n$ become infinite. Therefore we may write (4.1114) and (4.1115), and sufficiency follows as in Theorem 4.111.

The same proof establishes the following two theorems:

THEOREM 4.132. *In order that a real* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 4.133. *In order that a real* $(T)$ *may be such that*

$$\Omega(\{\sigma_n\}, A) \leqq \Omega(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

The five theorems 4.21–4.233 give necessary and sufficient conditions that $(T)$ shall transform sequences of zero oscillation into sequences of zero oscillation.

THEOREM 4.21. *In order that* $(T)$ *may be such that* $\Omega(\{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, *such that* $\Omega(\{s_n\}, A) = 0$, *no further conditions need be imposed upon* $a_{nk}$.

Letting $\{x_i\}$ be any sequence of points of $A$, and $q$ be an arbitrarily small positive number, we can choose an index $p$ for which (4.1112) holds; and using (4.1113) obtain (4.1116) precisely as in Theorem 4.111. But

$$\limsup_{m\to\infty,\ n\to\infty} (qB_mB_n) \leqq qB^2;$$

hence $\Omega(\{\sigma_n\}, A) \leqq qB^2$. Since $qB^2$ is an arbitrarily small positive number, $\Omega(\{\sigma_n\}, A) = 0$ and the theorem is proved.

THEOREM 4.22. *In order that* $(T)$ *may be such that* $\Omega(\{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, *such that* $\Omega(\{s_n\}, A) = 0$, *no further conditions need be imposed on* $a_{nk}$.

This theorem follows from the preceding, for if $\Omega(\{s_n\}, A) = 0$, $A$ being a finite set, then $s_n(x)$ is bounded over $A$ for all $n$.

THEOREM 4.231. *In order that* $(T)$ *may be such that* $\Omega(\{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that* $\Omega(\{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

Necessity follows from Lemma 4.02 (or 4.03). The sufficiency proof is a modification of that of Theorem 4.131 in the same sense that the proof of Theorem 4.21 is a modification of the sufficiency proof of Theorem 4.111. This proof also establishes the following two theorems:

THEOREM 4.232. *In order that a real* $(T)$ *may be such that* $\Omega(\{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that* $\Omega(\{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

THEOREM 4.233. *In order that a real* $(T)$ *may be such that* $\Omega(\{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, *such that* $\Omega(\{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

## 5. Continuous oscillation and convergence at a point

Let $\{f_n(x)\}$ be defined over a set $A$ and let $x_0$ be a point of $A^0$. Then the *continuous oscillation of* $\{f_n(x)\}$ *at* $x_0$ *over* $A$ is denoted by $\Omega(x_0; \{f_n\}, A)$ and is defined as follows: For each sequence $\{x_i\}$ of points of $A$ with the limit $x_0$, form

$$\limsup_{\substack{m\to\infty, n\to\infty \\ i\to\infty, j\to\infty}} |f_m(x_i) - f_n(x_j)| = v;$$

the least upper bound of all such $v$ is $\Omega(x_0; \{f_n\}, A)$. A sequence $\{f_n(x)\}$ is said to converge continuously at $x_0$ over $A$ when $\Omega(x_0; \{f_n\}, A) = 0$. We may consider $\Omega(x_0; \{f_n\}, A)$ as a measure of the deviation from continuous convergence of $\{f_n(x)\}$ at $x_0$ over $A$; hence the relation $\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$ signifies that $\{\sigma_n(x)\}$ is at least as nearly continuously convergent at $x_0$ over $A$ as $\{s_n(x)\}$ is. The continuous oscillation of a real sequence at a point is related to the ultimate bounds of the sequence at the point by the following lemma, the proof of which is similar to that of Lemma 4.06.

Lemma 5.01. *If* $\{f_n(x)\}$ *is a real sequence defined over a set* $A$ *such that* $x_0$ *is in* $A^0$, *then*

$$\Omega(x_0; \{f_n\}, A) = \Gamma(x_0; \{f_n\}, A) - \gamma(x_0; \{f_n\}, A)^*$$

*whenever the right member has a meaning; otherwise* $\Omega(x_0; \{f_n\}, A) = +\infty$.

Using this lemma, we may note that the theorems of this section which involve real transformations and sequences are related to the latter theorems of §2 as the corresponding theorems of §4 are related to the latter theorems of §1.

The eight theorems 5.111–5.133 give necessary and sufficient conditions that $(T)$ shall not increase the continuous oscillation of a sequence at a point.

Theorem 5.111. *In order that* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

Necessity follows from Lemma 4.01. To establish sufficiency, choose $M$ such that $|s_n(x)| < M$ over $D$ for all $n$, let $\{x_i\}$ be any sequence of points of $A$ with the limit $x_0$, and let $q$ be any number greater than $\Omega(x_0; \{s_n\}, A)$. Then there is an index $p$ for which (4.1112) holds; let $p$ be increased if necessary

* For real sequences $\Omega(x_0; \{f_n\}, A)$ is sometimes defined by this relation. Hahn, p. 261.

so that $x_i$ is in $D$ for $i>p$. Using (4.1113), we obtain (4.1117) exactly as in Theorem 4.111; therefore $\Omega(x_0; \{\sigma_n\}, A) \leqq q$ and sufficiency follows. The same proof establishes

THEOREM 5.112. *In order that a real* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

The next two theorems result from the preceding two, for if $\Omega(x_0; \{s_n\}, A)$ is finite and $x_0$ is an isolated point of $A$, then $s_n(x)$ is bounded over a neighborhood $D$ of $x_0$ in $A$ for all $n$.

THEOREM 5.121. *In order that* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point,* $C_4$ *is necessary and sufficient.*

THEOREM 5.122. *In order that a real* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point,* $C_4$ *is necessary and sufficient.*

From Theorems 5.112 and 5.122 we obtain

THEOREM 5.123. *In order that a real* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 5.131. *In order that* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

The proof of this theorem is a modification of that of Theorem 4.131 in the same sense that the proof of Theorem 5.111 is a modification of that of Theorem 4.111. The same proof establishes the next two theorems.

THEOREM 5.132. *In order that a real* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 5.133. *In order that* $(T)$ *may be such that*

$$\Omega(x_0; \{\sigma_n\}, A) \leqq \Omega(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

The five theorems 5.21–5.233, the proofs of which are modifications of the proofs of the corresponding theorems of §4, give necessary and sufficient conditions that $(T)$ shall transform sequences which converge continuously at a point into sequences which converge continuously at the point.

THEOREM 5.21. *In order that* $(T)$ *may be such that* $\Omega(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that* $\Omega(x_0; \{s_n\}, A) = 0$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 5.22. *In order that* $(T)$ *may be such that* $\Omega(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point, such that* $\Omega(x_0; \{s_n\}, A) = 0$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 5.231. *In order that* $(T)$ *may be such that* $\Omega(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that* $\Omega(x_0; \{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

THEOREM 5.232. *In order that a real* $(T)$ *may be such that* $\Omega(x_0; \{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that* $\Omega(x_0; \{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

THEOREM 5.233. *In order that a real* $(T)$ *may be such that* $\Omega(x_0; \{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that* $\Omega(x_0; \{s_n\}, A) = 0$, $C_5$ *is necessary and sufficient.*

## 6. CONTINUOUS OSCILLATION AND CONVERGENCE AT ALL POINTS AND LIMIT POINTS OF A SET

In comparing the theorems of this section with those of §4, the following facts, included as lemmas, should be noted.

LEMMA 6.01. *If* $\Omega(\{f_n\}, A) \leqq \Omega(\{F_n\}, A)$ *where* $A$ *contains more than one point, then it is not necessarily true that* $\Omega(x; \{f_n\}, A) \leqq \Omega(x; \{F_n\}, A)$ *over* $A$.

This lemma is established by considering the following simple sequences. Let $x_1$ and $x_2$ be two points of $A$ and define $f_n(x_1) = (-1)^n$, $f_n(x) = 0$ for $x \neq x_1$, and $F_n(x_2) = (-1)^n$, $F_n(x) = 0$ for $x \neq x_2$.

From the definitions of the functions involved we obtain at once

LEMMA 6.02. *If* $\Omega(\{f_n\}, A) = 0$, $A$ *being an arbitrary set, then* $\Omega(x; \{f_n\}, A) = 0$ *over* $A^0$.

LEMMA 6.03. *If* $\Omega(x; \{f_n\}, A) \leqq \Omega(x; \{F_n\}, A)$ *over* $A^0$ *and* $A$ *is compact, then* $\Omega(\{f_n\}, A) \leqq \Omega(\{F_n\}, A)$.

Suppose $A$ is compact and $\Omega(\{f_n\}, A) > \Omega(\{F_n\}, A)$; we shall show that the hypothesis is contradicted. Choose a sequence $\{x_i\}$ of points of $A$ and four sequences $\{m_\alpha\}$, $\{n_\alpha\}$, $\{i_\alpha\}$ and $\{j_\alpha\}$ of indices such that

$$\lim_{\alpha \to \infty} | f_{m_\alpha}(x_{i_\alpha}) - f_{n_\alpha}(x_{j_\alpha}) | = \Omega(\{f_n\}, A).$$

Since $A$ is compact, let $\{x_i'\}$ be a subsequence of $\{x_i\}$ with a limit, say $x_0$; then $\Omega(x_0; \{f_n\}, A) \geqq \Omega(\{f_n\}, A)$. It follows at once from the definitions of the functions involved that $\Omega(\{F_n\}, A) \geqq \Omega(x; \{F_n\}, A)$ over $A^0$ and we find on combining inequalities that $\Omega(x_0; \{f_n\}, A) > \Omega(x; \{F_n\}, A)$ over $A^0$; thus the hypothesis is contradicted and the lemma is proved. If in this lemma we set $F_n(x) = 0$ over $A$ for all $n$ we obtain

LEMMA 6.04. *If* $\Omega(x; \{f_n\}, A) = 0$ *over* $A^0$ *and* $A$ *is compact, then* $\Omega(\{f_n\}, A) = 0$.

The condition "$A$ is compact" cannot be removed* from Lemma 6.04 and it follows that it cannot be removed from Lemma 6.03.

The theorems of this section are obtained from those of §5 as theorems of §3 are obtained from §2.

The eight theorems 6.111–6.133 give necessary and sufficient conditions that $(T)$ shall not increase the continuous oscillations of sequences at the points and limit points of a set.

---

* To show this, let $\{x_n\}$ be a sequence of points of $A$ without limit points and define $f_n(x_n) = 1$ and $f_n(x) = 0$, $x \neq x_n$; then $\Omega(x; \{f_n\}, A) = 0$ over $A^0$, $\Omega(\{f_n\}, A) = 1$, and the lemma fails.

THEOREM 6.111. *In order that* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$, *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 6.112. *In order that a real* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 6.121. *In order that* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points,* $C_4$ *is necessary and sufficient.*

THEOREM 6.122. *In order that a real* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points,* $C_4$ *is necessary and sufficient.*

THEOREM 6.123. *In order that a real* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded above (below) over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 6.131. *In order that* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 6.132. *In order that a real* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point,* $C_4$ *and* $C_5$ *are necessary and sufficient.*

THEOREM 6.133. *In order that a real* $(T)$ *may be such that*

$$\Omega(x; \{\sigma_n\}, A) \leqq \Omega(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above (below) over* $A$ *for all* $n$, $C_4$ *and* $C_5$ *are necessary and sufficient.*

The five theorems 6.21–6.233 give necessary and sufficient conditions that $(T)$ shall transform sequences which converge continuously at each point and limit point of a set into sequences having the same property.

THEOREM 6.21. *In order that* $(T)$ *may be such that* $\Omega(x; \{\sigma_n\}, A) = 0$ *over* $A^0$ *for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded in the neighborhood of each point of* $A^0$ *for all* $n$, *such that* $\Omega(x; \{s_n\}, A) = 0$ *over* $A^0$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 6.22. *In order that* $(T)$ *may be such that* $\Omega(x; \{\sigma_n\}, A) = 0$ *over* $A$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points, such that* $\Omega(x; \{s_n\}, A) = 0$ *over* $A$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 6.231. *In order that* $(T)$ *may be such that* $\Omega(x; \{\sigma_n\}, A) = 0$ *over* $A^0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point such that* $\Omega(x; \{s_n\}, A) = 0$ *over* $A^0$, $C_5$ *is necessary and sufficient.*

THEOREM 6.232. *In order that a real* $(T)$ *may be such that* $\Omega(x; \{\sigma_n\}, A) = 0$ *over* $A^0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that* $\Omega(x; \{s_n\}, A) = 0$ *over* $A^0$, $C_5$ *is necessary and sufficient.*

THEOREM 6.233. *In order that a real* $(T)$ *may be such that* $\Omega(x; \{\sigma_n\}, A) = 0$ *over* $A^0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above (below) over* $A$ *for all* $n$, *such that* $\Omega(x; \{s_n\}, A) = 0$ *over* $A^0$, $C_5$ *is necessary and sufficient.*

## CHAPTER III. UNIFORM OSCILLATION AND CONVERGENCE

In this chapter the transformations and sequences may, except in cases where a specific statement to the contrary is made, be either real or complex.

We shall now prove together

LEMMAS 7.01, 7.02. *Let* $A$ *be an infinite set and let* $\{x_\alpha\}$ *be a sequence of distinct points of* $A$. *In case* $A$ *has a limit point* $x_0$, *let* $\{x_\alpha\}$ *have the limit* $x_0$. *If* $(T)$, *real or complex, fails to satisfy* $C_6$, *then there is a real sequence* $\{s_n(x)\}$ *bounded above (below) over* $A$ *such that*

$$\limsup_{m\to\infty, n\to\infty, i\to\infty} |\sigma_m(x_i) - \sigma_n(x_i)| > \limsup_{m\to\infty, n\to\infty, i\to\infty} |s_m(x_i') - s_n(x_i')| = 0$$

*where* $\{x_i'\}$ *is any sequence of points of* $A$.

From a denial of $C_6$ it follows that there is a sequence $\{n_\alpha\}$ such that

$$\lim_{\alpha\to\infty} n_\alpha = +\infty \text{ and } \sum_{k=1}^{n_\alpha} a_{n_\alpha,k} \neq 1 \text{ for } \alpha = 1,2,3,\cdots.$$

Define a real function $s(x)$ over $A$, bounded above (below) over $A$ such that

$$s(x_\alpha) = (-1)^h / \left|1 - \sum_{k=1}^{n_\alpha} a_{n_\alpha,k}\right|$$

where $h=1$ (2); and let $s_n(x)=s(x)$, $n=1, 2, 3, \cdots$. Then $s_n(x)$ is bounded above (below) over $A$ for all $n$; and $s_m(x)-s_n(x)=0$ over $A$ so that

$$\limsup_{m\to\infty,\, n\to\infty,\, i\to\infty} |\, s_m(x_i') - s_n(x_i')\,| = 0$$

where $\{x_i'\}$ is any sequence of points of $A$. But

$$\limsup_{m\to\infty,n\to\infty,i\to\infty} |\, \sigma_m(x_i) - \sigma_n(x_i)\,| = \limsup_{m\to\infty,n\to\infty,i\to\infty} \left( |\, s(x_i)\,| \left| \sum_{k=1}^{m} a_{mk} - \sum_{k=1}^{n} a_{nk} \right| \right)$$

$$\geqq \limsup_{n\to\infty,\, i\to\infty} \left( |\, s(x_i)\,| \left| 1 - \sum_{k=1}^{n} a_{nk} \right| \right) \geqq \limsup_{\alpha\to\infty} \left( |\, s(x_\alpha)\,| \left| 1 - \sum_{k=1}^{n_\alpha} a_{n_\alpha,k} \right| \right) = 1$$

and the lemmas are proved.

## 7. Uniform oscillation and convergence over a set

Let $\{f_n(x)\}$ be defined over any set $A$. Then the *uniform oscillation of the sequence $\{f_n(x)\}$ over the set $A$* is denoted by $O(\{f_n\}, A)$ and is defined as follows: For each sequence $\{x_i\}$ of points of $A$, form

$$\limsup_{m\to\infty,\, n\to\infty,\, i\to\infty} |\, f_m(x_i) - f_n(x_i)\,| = v;$$

the least upper bound of all such $v$ is $O(\{f_n\}, A)$. It is well known that a necessary and sufficient condition that a sequence $\{f_n(x)\}$ converge uniformly over a set $A$ is that $O(\{f_n\}, A)=0$. We may consider $O(\{f_n\}, A)$ as a measure of the deviation from uniform convergence of $\{f_n(x)\}$ over $A$; hence the relation $O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$ signifies that $\{\sigma_n(x)\}$ is at least as nearly uniformly convergent over $A$ as $\{s_n(x)\}$ is.

The eight theorems 7.111–7.133 give necessary and sufficient conditions that $(T)$ shall not increase uniform oscillations of sequences over a set.

Theorem 7.111. *In order that $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every sequence $\{s_n(x)\}$, defined over an arbitrary set $A$ and bounded over $A$ for all $n$, $C_4$ is necessary and sufficient.*

Necessity follows from Lemma 4.01. The sufficiency proof is verbatim that of Theorem 4.111 except that $j$ is replaced by $i$ and that the $\Omega$-function is replaced by the $O$-function. The same proof establishes

THEOREM 7.112. *In order that a real $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every real sequence $\{s_n(x)\}$, defined over an arbitrary set $A$ and bounded over $A$ for all $n$, $C_4$ is necessary and sufficient.*

Noting that if $O(\{s_n\}, A)$ is finite and $A$ is a finite set, then $s_n(x)$ is bounded over $A$ for all $n$, we obtain the next three theorems.

THEOREM 7.121. *In order that $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every sequence $\{s_n(x)\}$, defined over a finite set $A$, $C_4$ is necessary and sufficient.*

THEOREM 7.122. *In order that a real $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every real sequence $\{s_n(x)\}$, defined over a finite set $A$, $C_4$ is necessary and sufficient.*

THEOREM 7.123. *In order that a real $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every real sequence $\{s_n(x)\}$, defined over a finite set $A$ and bounded above (below) over $A$ for all $n$, $C_4$ is necessary and sufficient.*

THEOREM 7.131. *In order that $(T)$ may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every sequence $\{s_n(x)\}$, defined over an infinite set $A$, $C_4$, $C_5$ and $C_6$ are necessary and sufficient.*

The necessity of $C_4$ follows from Lemma 4.01; that of $C_5$ and $C_6$ from Lemmas 4.04 or 4.05 and 7.01 or 7.02 respectively since for the sequences there defined, $O(\{s_n\}, A)=0$ while $O(\{\sigma_n\}, A)>0$. If $O(\{s_n\}, A)=+\infty$, no proof of sufficiency is required. If $O(\{s_n\}, A)$ is finite, let $q$ be any greater number and let $\{x_i\}$ be any sequence of points of $A$; then there is an index $p$ such that $|s_\mu(x_i)-s_\nu(x_i)|<q$ for $\mu>p$, $\nu>p$, $i>p$. Using $C_5$ and $C_6$ we can choose an index $N$ such that $a_{nk}=0$, $k=1, 2, 3, \cdots, p$ for $n>N$ and

also $\sum_{k=p+1}^{n} a_{nk}=1$ for $n>N$. Then, considering the identity (4.1113) with $j$ replaced by $i$, the first four terms of the right member vanish for $m>N$, $n>N$; hence we obtain successively (4.1114), (4.1115), (4.1116), and, using $C_4$, (4.1117) with $j$ replaced by $i$. Therefore $O(\{\sigma_n\}, A)\leqq q$, $O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$ and the theorem is proved. The same proof establishes the next two theorems.

THEOREM 7.132. *In order that a real* $(T')$ *may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, $C_4$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 7.133. *In order that a real* $(T)$ *may be such that*

$$O(\{\sigma_n\}, A) \leqq O(\{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

The five theorems 7.21–7.233, the proofs of which are modifications of proofs of earlier theorems of this section, give necessary and sufficient conditions that $(T)$ shall transform sequences which converge uniformly over a set into sequences having the same property.

THEOREM 7.21. *In order that* $(T)$ *may be such that* $O(\{\sigma_n\}, A)=0$ *for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, *such that* $O(\{s_n\}, A)=0$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 7.22. *In order that* $(T)$ *may be such that* $O(\{\sigma_n\}, A)=0$ *for every sequence* $\{s_n(x)\}$, *defined over a finite set* $A$, *such that* $O(\{s_n\}, A)=0$, *no further conditions need be imposed on* $a_{nk}$.

THEOREM 7.231. *In order that* $(T)$ *may be such that* $O(\{\sigma_n\}, A)=0$ *for every sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that* $O(\{s_n\}, A)=0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 7.232. *In order that a real* $(T)$ *may be such that* $O(\{\sigma_n\}, A)=0$ *for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$, *such that* $O(\{s_n\}, A)=0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 7.233. *In order that a real* $(T)$ *may be such that* $O(\{\sigma_n\}, A)=0$ *for every real sequence* $\{s_n(x)\}$, *defined over an infinite set* $A$ *and bounded above (below) over* $A$ *for all* $n$, *such that* $O(\{s_n\}, A)=0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

## 8. Uniform oscillaton and convergence at a point

Let $\{f_n(x)\}$ be defined over a set $A$ and let $x_0$ be a point of $A^0$. Then the *uniform oscillation of* $\{f_n(x)\}$ *at* $x_0$ *over* $A$ (or measure of the deviation from uniform convergence of $\{f_n(x)\}$ *at* $x_0$ *over* $A$) is denoted by $O(x_0; \{f_n\}, A)$ and is defined as follows: For each sequence $\{x_i\}$ of points of $A$ with the limit $x_0$, form

$$\limsup_{m\to\infty, n\to\infty, i\to\infty} |f_m(x_i) - f_n(x_i)| = v;$$

the least upper bound of all such $v$ is $O(x_0; \{f_n\}, A)$. A sequence $\{f_n(x)\}$ is said to converge uniformly at $x_0$ over $A$ when $O(x_0; \{f_n\}, A) = 0$. The proofs of the theorems of this section are similar to the proofs of the corresponding theorems of §5, the essential difference occurring where the proofs of §5 refer to §4; in such cases the proofs of this section refer to §7.

The eight theorems 8.111–8.133 give necessary and sufficient conditions that $(T)$ shall not increase uniform oscillations of sequences at a point.

Theorem 8.111. *In order that* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

Theorem 8.112. *In order that a real* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

Theorem 8.121. *In order that* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point,* $C_4$ *is necessary and sufficient.*

Theorem 8.122. *In order that a real* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point,* $C_4$ *is necessary and sufficient.*

THEOREM 8.123. *In order that a real* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 8.131. *In order that* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point,* $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

THEOREM 8.132. *In order that a real* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point,* $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

THEOREM 8.133. *In order that a real* $(T)$ *may be such that*

$$O(x_0; \{\sigma_n\}, A) \leqq O(x_0; \{s_n\}, A)$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

The five theorems 8.21–8.233 give necessary and sufficient conditions that $(T)$ shall transform sequences which converge uniformly at a point into sequences having the same property.

THEOREM 8.21. *In order that* $(T)$ *may be such that* $O(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *such that* $x_0$ *is in* $A^0$ *and bounded over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that* $O(x_0; \{s_n\}, A) = 0$, *no further conditions need be imposed upon* $a_{nk}$.

THEOREM 8.22. *In order that* $(T)$ *may be such that* $O(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is an isolated point, such that* $O(x_0; \{s_n\}, A) = 0$, *no further conditions need be imposed upon* $a_{nk}$.

THEOREM 8.231. *In order that* $(T)$ *may be such that* $O(x_0; \{\sigma_n\}, A) = 0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that* $O(x_0; \{s_n\}, A) = 0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 8.232. *In order that a real* $(T)$ *may be such that* $O(x_0; \{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point, such that* $O(x_0; \{s_n\}, A) = 0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 8.233. *In order that a real* $(T)$ *may be such that* $O(x_0; \{\sigma_n\}, A) = 0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *of which* $x_0$ *is a limit point and bounded above (below) over a neighborhood* $D$ *of* $x_0$ *in* $A$ *for all* $n$, *such that* $O(x_0; \{s_n\}, A) = 0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

## 9. UNIFORM OSCILLATION AND CONVERGENCE AT ALL POINTS AND LIMIT POINTS OF A SET

In comparing the theorems of this section with those of §7, the following facts, given as lemmas, the proofs of which are similar to the proofs of the corresponding lemmas of §6, should be noted.

LEMMA 9.01. *If* $O(\{f_n\}, A) \leqq O(\{F_n\}, A)$ *where* $A$ *contains more than one point, then it is not necessarily true that* $O(x; \{f_n\}, A) \leqq O(x; \{F_n\}, A)$ *over* $A$.

LEMMA 9.02. *If* $O(\{f_n\}, A) = 0$, $A$ *being an arbitrary set, then* $O(x; \{f_n\}, A) = 0$ *over* $A^0$.

LEMMA 9.03. *If* $O(x; \{f_n\}, A) \leqq O(x; \{F_n\}, A)$ *over* $A^0$ *and* $A$ *is compact, then* $O(\{f_n\}, A) \leqq O(\{F_n\}, A)$.

LEMMA 9.04. *If* $O(x; \{f_n\}, A) = 0$ *over* $A^0$ *and* $A$ *is compact, then* $O(\{f_n\}, A) = 0$.

The theorems of this section are obtained from those of §8 as the theorems of §6 were obtained from those of §5. The eight theorems 9.111–9.133 give necessary and sufficient conditions that $(T)$ shall not increase the uniform oscillations of sequences at the points and limit points of a set.

THEOREM 9.111. *In order that* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 9.112. *In order that a real* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 9.121. *In order that* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points,* $C_4$ *is necessary and sufficient.*

THEOREM 9.122. *In order that a real* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points,* $C_4$ *is necessary and sufficient.*

THEOREM 9.123. *In order that a real* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points and bounded above (below) over* $A$ *for all* $n$, $C_4$ *is necessary and sufficient.*

THEOREM 9.131. *In order that* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A^0$$

*for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point,* $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

THEOREM 9.132. *In order that a real* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point,* $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

THEOREM 9.133. *In order that a real* $(T)$ *may be such that*

$$O(x; \{\sigma_n\}, A) \leqq O(x; \{s_n\}, A) \text{ over } A^0$$

*for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above (below) over* $A$ *for all* $n$, $C_4$, $C_5$, *and* $C_6$ *are necessary and sufficient.*

The five theorems 9.21–9.233 give necessary and sufficient conditions that $(T)$ shall transform sequences which converge uniformly at each point and limit point of a set into sequences having the same property.

THEOREM 9.21. *In order that* $(T)$ *may be such that* $O(x; \{\sigma_n\}, A)=0$ *over* $A^0$ *for every sequence* $\{s_n(x)\}$, *defined over an arbitrary set* $A$ *and bounded in the neighborhood of each point of* $A^0$ *for all* $n$, *such that* $O(x; \{s_n\}, A)=0$ *over* $A^0$, *no further conditions need be imposed upon* $a_{nk}$.

THEOREM 9.22. *In order that* $(T)$ *may be such that* $O(x; \{\sigma_n\}, A)=0$ *over* $A$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *without limit points, such that* $O(x; \{s_n\}, A)=0$ *over* $A$, *no further conditions need be imposed upon* $a_{nk}$.

THEOREM 9.231. *In order that* $(T)$ *may be such that* $O(x; \{\sigma_n\}, A)=0$ *over* $A^0$ *for every sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that* $O(x; \{s_n\}, A)=0$ *over* $A^0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 9.232. *In order that a real* $(T)$ *may be such that* $O(x; \{\sigma_n\}, A)=0$ *over* $A^0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point, such that* $O(x; \{s_n\}, A)=0$ *over* $A^0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

THEOREM 9.233. *In order that a real* $(T)$ *may be such that* $O(x; \{\sigma_n\}, A)=0$ *over* $A^0$ *for every real sequence* $\{s_n(x)\}$, *defined over a set* $A$ *with at least one limit point and bounded above (below) over* $A$ *for all* $n$, *such that* $O(x; \{s_n\}, A)=0$ *over* $A^0$, $C_5$ *and* $C_6$ *are necessary and sufficient.*

CORNELL UNIVERSITY,
ITHACA, N. Y.

# SOME DIFFERENTIAL AND ALGEBRAIC CONSEQUENCES OF THE EINSTEIN FIELD EQUATIONS*

BY

K. W. LAMSON

**Introduction.** If a set of four directions in a Riemannian four-dimensional space, $V_4$, is orthogonal, then the $ds^2$ can be expressed in terms of their sixteen parameters, $h_k^{\alpha}(x_0, x_1, x_2, x_3)$, as in Einstein's recent papers.

The first purpose of this paper is to set up sixteen invariant linear first-order partial differential equations in these parameters (§2). The solutions of these equations include all solutions for empty space of the Einstein field equations of 1917. There is a restriction which excludes some special cases. In addition to the $h_k^{\alpha}$, these sixteen equations contain four linear combinations of the components of the curvature tensor. These four combinations are to be taken as independent variables, $x^k$.

Since only alternating tensors appear it is convenient to use Cartan's notation† for symbolic differential forms and for their derivatives and products. Covariant differentiation in the sense of the absolute differential calculus is not used, except in §4.

The components of the curvature tensor may be taken as coefficients in the equation of a quadratic line complex in a three-dimensional projective space, $P_3$‡§. The directions $h_k^{\alpha}$ correspond to the vertices of a tetrahedron in $P_3$. Bivector and simple bivector correspond to linear complex and special complex. Where there is no danger of misunderstanding, the language of $V_4$ will be used interchangeably with that of $P_3$. The second purpose of this paper is the application of some of the theory of quadratic complexes to the study of the curvatures in $V_4$.¶

The lines which lie in a plane in $P_3$ and which belong to the quadratic complex are tangent to a conic. The envelope of planes for which this conic

---

* Presented to the Society, December 27, 1923, and September 12, 1930; received by the editors February 20, 1930.

† Goursat, *Leçons sur le Problème de Pfaff*, chapters I and III.

‡ D. J. Struik, *On sets of principal directions in a Riemannian manifold in four dimensions*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 7 (1928), p. 193.

§ E. Kretschmann, Annalen der Physik, vol. 53 (1917), p. 592.

¶ Jessop, *Treatise on the Line Complex*. See also Hudson, *Kummer's Quartic Surface*, and Zindler's article in the Encyclopedia, IIIC8.

becomes two plane pencils is a Kummer surface. For the tropes of this surface the two vertices of these pencils coincide in a point which lies on the surface. Each of the sixteen tropes is tangent to the surface along a conic.

Quadratic complexes were classified by Klein in types characterized by the elementary divisors of their quadratic forms taken with that of the Plücker identity. In this paper the type [111111] is assumed throughout.

In §6, $\theta$-functions are introduced as a further application of the theory of complexes, and also for reasons of symmetry. No one of the ten quadrics or of the many tetrahedra associated with the curvature tensor is held out against the others.

1. **Notation.** The differential of length is the sum of the squares of four Pfaff expressions,

$$\omega_k = h_{k|\alpha} dx^\alpha \qquad (k = 0, 1, 2, 3),$$
$$ds^2 = \omega_0^2 + \omega_1^2 + \omega_2^2 + \omega_3^2 .$$

The Ricci coefficients of rotation are given by

$$2\gamma_{ijk} = h_{k|\alpha}(h_{i\rho}^{\alpha} h_j^{\rho} - h_{j\rho}^{\alpha} h_i^{\rho}) + h_{j|\alpha}(h_{i\rho}^{\alpha} h_k^{\rho} - h_{k\rho}^{\alpha} h_i^{\rho}) + h_{i|\alpha}(h_{k\rho}^{\alpha} h_j^{\rho} - h_{j\rho}^{\alpha} h_k^{\rho}),$$

where the matrix $h_k^\alpha$ is the reciprocal of $h_{k|\alpha}$, and the subscript $\rho$ denotes ordinary differentiation.

The components of the curvature tensor are $\gamma_{ij,kl}$, referred to the directions $h_k^\alpha$, found by putting three of the $\omega_k$ equal to zero.* Greek and Roman indices are used when the axis-system is given by $dx^\alpha = 0$ and $\omega_k = 0$, respectively.

The symbolic product of two Pfaffians is illustrated by

$$[\omega_1 \omega_2] = h_{1|\alpha} h_{2|\beta} (dx^\alpha \delta x^\beta - dx^\beta \delta x^\alpha).$$

Symbolic forms of the second and third degree are linear combinations of

$$dx^\alpha \delta x^\beta - dx^\beta \delta x^\alpha$$

or

$$dx^\alpha \delta x^\beta \Delta x^\gamma + dx^\beta \delta x^\gamma \Delta x^\alpha + dx^\gamma \delta x^\alpha \Delta x^\beta - dx^\alpha \delta x^\gamma \Delta x^\beta - dx^\beta \delta x^\alpha \Delta x^\gamma - dx^\gamma \delta x^\beta \Delta x^\alpha .$$

If two forms have a common factor of the first degree their product is zero. When the square of a form is written, ordinary multiplication is meant.

---

* Levi-Civita, *A Simplified Presentation of Einstein's Unified Field Equations*, p. 6, formula 8; *Vereinfachte Herstellung der Einstein'schen einheitlichen Feldgleichungen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1929, p. 139, formula (8).

The derivative of a form of degree $n$ is a form of degree $n+1$, is indicated by a prime, and illustrated by

$$\omega_k' = \gamma_{kji}[\omega_i\omega_j].$$

In $V_4$ the $\Omega_k$ are defined as follows in terms of symbolic products of the $\omega_k$. In $P_3$ the definitions in terms of the Plücker line coördinates are used:

$$\begin{aligned}
\Omega_1 &= [\omega_0\omega_1] + [\omega_2\omega_3] = p_{01} + p_{23},\\
\Omega_2 &= [\omega_0\omega_2] + [\omega_3\omega_1] = p_{02} + p_{31},\\
\Omega_3 &= [\omega_0\omega_3] + [\omega_1\omega_2] = p_{03} + p_{12};
\end{aligned} \tag{1.1}$$

$$\begin{aligned}
\bar{\Omega}_1 &= -[\omega_0\omega_1] + [\omega_2\omega_3] = -p_{01} + p_{23},\\
\bar{\Omega}_2 &= -[\omega_0\omega_2] + [\omega_3\omega_1] = -p_{02} + p_{31},\\
\bar{\Omega}_3 &= -[\omega_0\omega_3] + [\omega_1\omega_2] = -p_{03} + p_{12}.
\end{aligned}$$

These expressions equated to zero are the six fundamental complexes of Klein.

The following equations define the Pfaff expressions $\mu_k$ and $\bar{\mu}_k$:

$$\begin{aligned}
\mu_1 &= (\gamma_{01k} + \gamma_{23k})\omega_k, & \bar{\mu}_1 &= (-\gamma_{01k} + \gamma_{23k})\omega_k,\\
\mu_2 &= (\gamma_{02k} + \gamma_{31k})\omega_k, & \bar{\mu}_2 &= (-\gamma_{02k} + \gamma_{31k})\omega_k,\\
\mu_3 &= (\gamma_{03k} + \gamma_{12k})\omega_k, & \bar{\mu}_3 &= (-\gamma_{03k} + \gamma_{12k})\omega_k.
\end{aligned} \tag{1.2}$$

Then, from (1.1) and (1.2),

$$\begin{aligned}
\Omega_1' &= [\Omega_2\mu_3] - [\Omega_3\mu_2], & \bar{\Omega}_1' &= [\bar{\Omega}_2\bar{\mu}_3] - [\bar{\Omega}_3\bar{\mu}_2],\\
\Omega_2' &= [\Omega_3\mu_1] - [\Omega_1\mu_3], & \bar{\Omega}_2' &= [\bar{\Omega}_3\bar{\mu}_1] - [\bar{\Omega}_1\bar{\mu}_3],\\
\Omega_3' &= [\Omega_1\mu_2] - [\Omega_2\mu_1], & \bar{\Omega}_3' &= [\bar{\Omega}_1\bar{\mu}_2] - [\bar{\Omega}_2\bar{\mu}_1].
\end{aligned} \tag{1.3}$$

To the hypercone of minimal directions in $V_4$ corresponds a quadric surface in $P_3$, which will be referred to as $ds^2$.

2. **First-order differential equations.** For the $ds^2$ of §1 the Einstein field equations for empty space can be written as follows:

$$\gamma_{01,01} - \gamma_{23,23} = 0, \tag{2.1}$$

$$\gamma_{02,02} - \gamma_{31,31} = 0, \tag{2.2}$$

$$\gamma_{03,03} - \gamma_{12,12} = 0, \tag{2.3}$$

$$\gamma_{02,03} + \gamma_{31,03} - \gamma_{02,12} - \gamma_{31,12} = 0, \tag{2.4}$$

$$\gamma_{02,03} - \gamma_{31,03} + \gamma_{02,12} - \gamma_{31,12} = 0, \tag{2.5}$$

$$\gamma_{03,01} + \gamma_{12,01} - \gamma_{03,23} - \gamma_{12,23} = 0, \tag{2.6}$$

$$\gamma_{03,01} - \gamma_{12,01} + \gamma_{03,23} - \gamma_{12,23} = 0, \tag{2.7}$$

$$\gamma_{01,02} + \gamma_{23,02} - \gamma_{01,31} - \gamma_{23,31} = 0, \tag{2.8}$$

$$\gamma_{01,02} - \gamma_{23,02} + \gamma_{01,31} - \gamma_{23,31} = 0. \tag{2.9}$$

A consequence of these nine is the tenth equation

$$\gamma_{01,01} + \gamma_{02,02} + \gamma_{03,03} = \text{const. (the cosmological } \lambda). \tag{2.10}$$

The nine equations can be written as the following three equations in symbolic differential forms of the second degree, each equivalent to six scalar equations:*

$$\mu_1' + [\mu_2\mu_3] = A_{11}\Omega_1 + A_{12}\Omega_2 + A_{13}\Omega_3, \tag{2.11}$$

$$\mu_2' + [\mu_3\mu_1] = A_{21}\Omega_1 + A_{22}\Omega_2 + A_{23}\Omega_3, \tag{2.12}$$

$$\mu_3' + [\mu_1\mu_2] = A_{31}\Omega_1 + A_{32}\Omega_2 + A_{33}\Omega_3 \qquad (A_{jk} = A_{kj}). \tag{2.13}$$

Symbolic multiplication by $\overline{\Omega}_1$, $\overline{\Omega}_2$, $\overline{\Omega}_3$ gives the equations (2.1) to (2.9) as indicated in the following table.

| | (2.11) | (2.12) | (2.13) |
|---|---|---|---|
| $\overline{\Omega}_1$ | (2.1) | (2.9) | (2.6) |
| $\overline{\Omega}_2$ | (2.8) | (2.2) | (2.5) |
| $\overline{\Omega}_3$ | (2.7) | (2.4) | (2.3) |

The $A_{jk}$ are undetermined coefficients which drop out for these multiplications, because the complexes $\Omega$ are in involution, or in other words, because the symbolic product of any two different ones is zero.

Another way of writing the field equations is as follows:

$$\bar{\mu}_1' + [\bar{\mu}_2\bar{\mu}_3] = \overline{A}_{11}\overline{\Omega}_1 + \overline{A}_{12}\overline{\Omega}_2 + \overline{A}_{13}\overline{\Omega}_3, \tag{2.14}$$

$$\bar{\mu}_2' + [\bar{\mu}_3\bar{\mu}_1] = \overline{A}_{21}\overline{\Omega}_1 + \overline{A}_{22}\overline{\Omega}_2 + \overline{A}_{23}\overline{\Omega}_3, \tag{2.15}$$

$$\bar{\mu}_3' + [\bar{\mu}_1\bar{\mu}_2] = \overline{A}_{31}\overline{\Omega}_1 + \overline{A}_{32}\overline{\Omega}_2 + \overline{A}_{33}\overline{\Omega}_3 \qquad (\overline{A}_{jk} = \overline{A}_{kj}). \tag{2.16}$$

The quadratic complex whose coefficients are the components of the curvature tensor is then

$$A_{11}\Omega_1^2 + A_{22}\Omega_2^2 + A_{33}\Omega_3^2 + \overline{A}_{11}\overline{\Omega}_1^2 + \overline{A}_{22}\overline{\Omega}_2^2 + \overline{A}_{33}\overline{\Omega}_3^2 + 2(A_{23}\Omega_2\Omega_3 + A_{31}\Omega_3\Omega_1 + A_{12}\Omega_1\Omega_2 + \overline{A}_{23}\overline{\Omega}_2\overline{\Omega}_3 + \overline{A}_{31}\overline{\Omega}_3\overline{\Omega}_1 + \overline{A}_{12}\overline{\Omega}_1\overline{\Omega}_2) = 0. \tag{2.17}$$

The products are ordinary, not symbolic.

If the complex is of type [111111], it has associated with it six fundamen-

* Cartan, *Equations de la Gravitation d'Einstein*, Journal de Mathématiques, (9), vol. 1 (1922), p. 153.

tal linear complexes.* From the way in which these are found it follows from (2.17) that *the field equations for empty space are the necessary and sufficient condition that the three linear complexes of one regulus of $ds^2$ be linear combinations of three of the fundamental complexes, and that the complexes of the other regulus be combinations of the remaining three.* If the complex is not of the above type this statement may require modification.

By multiplying (2.11) symbolically with $\Omega_1$, and (2.12) with $\Omega_3$ (or (2.13) with $\Omega_2$) it is seen that

$$2A_{11} = \gamma_{01,01} + 2\gamma_{01,23} + \gamma_{23,23} \quad \text{(cyclic 1, 2, 3)},$$
$$2A_{23} = \gamma_{02,03} + \gamma_{31,03} + \gamma_{02,12} + \gamma_{31,12} \quad \text{(cyclic 1, 2, 3)},$$

and similarly

$$2\overline{A}_{11} = \gamma_{01,01} - 2\gamma_{01,23} + \gamma_{23,23} \quad \text{(cyclic 1, 2, 3)},$$
$$2\overline{A}_{23} = \gamma_{02,03} - \gamma_{31,03} - \gamma_{02,12} + \gamma_{31,12} \quad \text{(cyclic 1, 2, 3)}.$$

So far there are nine independent equations in the sixteen unknowns, $h_k^\alpha$. For the purposes of this section there are added the six equations

$$A_{jk} = \overline{A}_{jk} = 0 \qquad (j \neq k). \tag{2.18}$$

This reduces the quadratic complex to the Klein form, and amounts to taking the axes of reference in $V_4$ as directions which are cut by the $P_3$ in any one of those six fundamental tetrahedra which are self-polar with respect to $ds^2$. In fact *the field equations in the case* [111111] *are satisfied if and only if there are six tetrahedra which simultaneously reduce the $ds^2$ to a sum of squares of Pfaffians and the quadratic complex to the Klein form.*

Differentiating (2.11), in the sense of Cartan, and using (2.18), we get

$$[\mu_2'\mu_3] - [\mu_3'\mu_2] = A_{11}\Omega_1' + [A_{11}'\Omega_1].$$

Substituting for $\mu_2'$, $\mu_3'$, and $\Omega_1'$ from (2.12), (2.13), and (1.3), we get the four scalar equations included in the third-degree symbolic equation

$$A_{22}[\Omega_2\mu_3] - A_{33}[\Omega_3\mu_2] = A_{11}([\Omega_2\mu_3] - [\Omega_3\mu_2]) + [A_{11}'\Omega_2],$$

and also get five other sets of four equations from (2.12) to (2.16).

These equations written in full are as follows:

$$\begin{aligned}
h_0^\alpha \partial A_{11}/\partial x^\alpha &= (A_{33} - A_{11})(\gamma_{022} + \gamma_{312}) - (A_{11} - A_{22})(\gamma_{033} + \gamma_{123}),\\
h_1^\alpha \partial A_{11}/\partial x^\alpha &= (A_{33} - A_{11})(-\gamma_{023} - \gamma_{313}) - (A_{11} - A_{22})(\gamma_{032} + \gamma_{122}),\\
h_2^\alpha \partial A_{11}/\partial x^\alpha &= (A_{33} - A_{11})(-\gamma_{020} - \gamma_{310}) - (A_{11} - A_{22})(-\gamma_{031} - \gamma_{121}),\\
h_3^\alpha \partial A_{11}/\partial x^\alpha &= (A_{33} - A_{11})(\gamma_{021} + \gamma_{311}) - (A_{11} - A_{22})(-\gamma_{030} - \gamma_{120});
\end{aligned} \tag{2.19}$$

* Jessop, loc. cit., p. 189, and Klein, Mathematische Annalen, vol. 2 (1870), p. 198.

(2.20) In (2.19) the indices (1, 2, 3) are replaced with (2, 3, 1);

(2.21) In (2.19) the indices (1, 2, 3) are replaced with (3, 1, 2);

$$
\begin{aligned}
h_0{}^\alpha \partial \overline{A}_{11}/\partial x^\alpha &= (\overline{A}_{33} - \overline{A}_{11})(\gamma_{022} - \gamma_{312}) - (\overline{A}_{11} - \overline{A}_{22})(\gamma_{033} - \gamma_{123}),\\
h_1{}^\alpha \partial \overline{A}_{11}/\partial x^\alpha &= (\overline{A}_{33} - \overline{A}_{11})(\gamma_{023} - \gamma_{313}) - (\overline{A}_{11} - \overline{A}_{22})(-\gamma_{032} + \gamma_{122}),\\
h_2{}^\alpha \partial \overline{A}_{11}/\partial x^\alpha &= (\overline{A}_{33} - \overline{A}_{11})(-\gamma_{020} + \gamma_{310}) - (\overline{A}_{11} - \overline{A}_{22})(\gamma_{031} - \gamma_{121}),\\
h_3{}^\alpha \partial \overline{A}_{11}/\partial x^\alpha &= (\overline{A}_{33} - \overline{A}_{11})(-\gamma_{021} + \gamma_{311}) - (\overline{A}_{11} - \overline{A}_{22})(-\gamma_{030} + \gamma_{120});
\end{aligned}
\tag{2.22}
$$

(2.23) In (2.21) the indices (1, 2, 3) are replaced with (2, 3, 1);

(2.24) In (2.21) the indices (1, 2, 3) are replaced with (3, 1, 2).

The addition of (2.19), (2.20), and (2.21) shows that

$$A_{11} + A_{22} + A_{33} = \text{const. (the cosmological } \lambda),$$

and similarly

$$\overline{A}_{11} + \overline{A}_{22} + \overline{A}_{33} = \lambda.$$

There remain sixteen independent equations, for example (2.20), (2.21), (2.23), and (2.24). If four of the $A$'s are independent, which is to be expected in the general case, they may be taken as independent variables. The main result of this section may be expressed as follows.

*Every solution of the Einstein field equations in empty space, of type* [111111], *and with four independent components of the curvature, is included (up to change of variable) among the solutions of the linear first-order equations* (2.19) *to* (2.24), *where*

$$
\begin{aligned}
A_{11} &= \lambda - x^0 - x^1, & \overline{A}_{11} &= \lambda - x^2 - x^3,\\
A_{22} &= x^0, & \overline{A}_{22} &= x^2,\\
A_{33} &= x^1, & \overline{A}_{33} &= x^3,
\end{aligned}
$$

*and the* $\gamma_{ijk}$ *are the Ricci coefficients of rotation.*

If there are any solutions of this type for which the $A$'s are not independent, the elimination of the $A$'s would lead to algebraic equations in the coefficients of rotation.

Conversely it can be shown by differentiation that a consequence of (2.19) etc. is

$$A_{jk} = \overline{A}_{jk} = 0 \qquad (j \neq k).$$

3. **Characteristic properties of the complex for the case $\lambda = 0$.** As before, it is possible to impose six conditions on the twelve $A$'s appearing in (2.11)

to (2.16). Suppose that this is done so that they are given by the following combinations of six new functions $b_1$, $b_2$, $b_3$, $a_{11}$, $a_{22}$, $a_{33}$:

$$A_{jk} = 2b_jb_k + a_{jk} \qquad (j, k = 1, 2, 3);$$
$$\overline{A}_{jk} = a_{jk} \qquad (a_{jk} = 0, j \neq k).$$

The quadratic complex is then referred to a Rosenhain tetrahedron of nodes and tropes, found by taking the polar planes of any node in $\overline{\Omega}_1$, $\overline{\Omega}_2$, $\overline{\Omega}_3$, and the poles of these planes with respect to the same complexes.*

Its equation in Plücker coördinates is

$$[b_1(p_{01} + p_{23}) + b_2(p_{02} + p_{31}) + b_3(p_{03} + p_{12})]^2 + 2a_{11}p_{01}p_{23} + 2a_{22}p_{02}p_{31} + 2a_{33}p_{03}p_{12} = \gamma_{ij,kl}p_{ij}p_{kl} = 0, \tag{3.1}$$

being given by adding the squared equation of a linear complex to that of a tetrahedral complex. If $\lambda = 0$, it follows from (2.10) that

$$b_1^2 + b_2^2 + b_3^2 = 0. \tag{3.2}$$

Therefore the linear complex is special (the bivector is simple) and its directrix is one of the lines of that regulus of $ds^2$ which is given by

$$\overline{\Omega}_1 = \overline{\Omega}_2 = \overline{\Omega}_3 = 0.$$

In Hudson's notation (p. 78) the vertices of the tetrahedron chosen above are the nodes corresponding to four dots in a row. Four dots in a column might have been used. This would lead to a line in the other regulus.

*It is characteristic of the case* $\lambda = 0$ *that the eight linear complexes associated with the eight Rosenhain tetrahedra of nodes self-polar with respect to* $ds^2$ *are special. Their lines are rulings of* $ds^2$*, four to each regulus.*

The rest of this section is devoted to formulas for these lines. There is a principal tangent on the Kummer surface whose tangent lines are second-order singular lines of the complex.† Let $(r_0, r_1, r_2, r_3)$ be the coördinates, referred to a fundamental tetrahedron, of the one point not a node where this curve touches the conic in one of the tropes $(t_0, t_1, t_2, t_3)$. Then $(r)$ is also the plane tangent to the same principal tangent through the node which is the pole of that trope in $ds^2$. For the node $(t)$ the complex cone degenerates into the pencil of lines in the plane $(r)$ through the point $(t)$. For the trope $(t)$ the complex conic degenerates into the pencil in the plane $(t)$ through the point $(r)$.

The four nodes (tropes) of the Rosenhain tetrahedron and the planes

* Jessop, loc. cit., p. 157, and Hudson, loc. cit., p. 83.

† Jessop, loc. cit., p. 95.

(points) of the corresponding pencils as just described are given in the following table.

| Nodes (tropes) referred to a tetrahedron which is | | | | | | | | Corresponding planes (points) referred to a tetrahedron which is | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fundamental | | | | Rosenhain | | | | Fundamental | | | | Rosenhain | | | |
| $t_0$, | $t_1$, | $t_2$, | $t_3$ | 1, | 0, | 0, | 0 | $r_0$, | $r_1$, | $r_2$, | $r_3$ | 0, | $b_1$, | $b_2$, | $b_3$ |
| $-t_1$, | $t_0$, | $t_3$, | $-t_2$ | 0, | 1, | 0, | 0 | $-r_1$, | $r_0$, | $r_3$, | $-r_2$ | $-b_1$, | 0, | $b_3$, | $-b_2$ |
| $-t_2$, | $-t_3$, | $t_0$, | $t_1$ | 0, | 0, | 1, | 0 | $-r_2$, | $-r_3$, | $r_0$, | $r_1$ | $-b_2$, | $-b_3$, | 0, | $b_1$ |
| $-t_3$, | $t_2$, | $-t_1$, | $t_0$ | 0, | 0, | 0, | 1 | $-r_3$, | $r_2$, | $-r_1$, | $r_0$ | $-b_3$, | $b_2$, | $-b_1$, | 0 |

From the usual formulas for the incidence of line and plane and from (3.2) it is seen that the four planes of the last set contain the directrix of the special complex ($b$). Therefore the determinant of the third set is zero and of rank two. The resulting equation is

$$r_0^2 + r_1^2 + r_2^2 + r_3^2 = 0 \tag{3.3}$$

which states that the point (plane) ($r$) lies on (is tangent to) the quadric $ds^2$. In the following table the four points in any row or column lie on a line. These eight lines are those referred to in the above italicized statement.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_0$, | $r_1$, | $r_2$, | $r_3$ | $-r_1$, | $r_0$, | $r_3$, | $-r_2$ | $-r_2$, | $-r_3$, | $r_0$, | $r_1$ | $-r_3$, | $r_2$, | $-r_1$, | $r_0$ |
| $r_1$, | $-r_0$, | $r_3$, | $-r_2$ | $r_0$, | $r_1$, | $-r_2$, | $-r_3$ | $-r_3$, | $r_2$, | $r_1$, | $-r_0$ | $r_2$, | $r_3$, | $r_0$, | $r_1$ |
| $r_2$, | $-r_3$, | $-r_0$, | $r_1$ | $r_3$, | $r_2$, | $r_1$, | $r_0$ | $r_0$, | $-r_1$, | $r_2$, | $-r_3$ | $-r_1$, | $-r_0$, | $r_3$, | $r_2$ |
| $r_3$, | $r_2$, | $-r_1$, | $-r_0$ | $-r_2$, | $r_3$, | $-r_0$, | $r_1$ | $r_1$, | $r_0$, | $r_3$, | $r_2$ | $r_0$, | $-r_1$, | $-r_2$, | $r_3$ |

In computing this result it is convenient to have the $\gamma_{ij,kl}$ or $A_{jk}$ expressed in terms of the $r_k$ and $t_k$. Let the coefficient of proportionality of the $t_k$ be chosen so that

$$\sum t_k^2 = 1.$$

Then the required expressions are obtained from the analytic statement that the complex cone through ($t$) is the plane ($r$) taken twice. All the $\gamma_{ij,kl}$ can be obtained from the following, which also give the factor of proportionality of the $r_k$:

$$\gamma_{01,01} + \gamma_{02,31} - \gamma_{03,12} = -(r_0r_1 + r_2r_3)/(t_0t_1 + t_2t_3) \quad \text{(cyclic 1, 2, 3);} \tag{3.4}$$

$$\gamma_{01,01} - \gamma_{02,31} + \gamma_{03,12} = -(-r_0r_1 + r_2r_3)/(-t_0t_1 + t_2t_3) \quad \text{(cyclic 1, 2, 3);} \tag{3.5}$$

$$\begin{aligned} r_0^2 &= \gamma_{01,01}t_1^2 + \gamma_{02,02}t_2^2 + \gamma_{03,03}t_3^2, \\ r_1^2 &= \gamma_{01,01}t_0^2 + \gamma_{02,02}t_3^2 + \gamma_{03,03}t_2^2, \\ r_2^2 &= \gamma_{01,01}t_3^2 + \gamma_{02,02}t_0^2 + \gamma_{03,03}t_1^2, \\ r_3^2 &= \gamma_{01,01}t_2^2 + \gamma_{02,02}t_1^2 + \gamma_{03,03}t_0^2. \end{aligned} \tag{3.6}$$

No use was made of $\lambda=0$ in the derivation of equations (3.4), (3.5), and (3.6), and they are true if $\lambda\neq 0$.

The addition of the four equations (3.6) and substitution from (2.10) give

$$r_0^2 + r_1^2 + r_2^2 + r_3^2 = \lambda. \tag{3.7}$$

The complex cone for each of the sixteen points $(r)$ in the tropes as described above, degenerates into the plane $(t)$ which contains that $(r)$, and another plane $(w)$. The coördinates of $(w)$ are readily obtained from its definition. Then the differences between the cases $\lambda=0$ and $\lambda\neq 0$ are given in the following table of incidences. The entries "yes" and "no" indicate whether there is incidence or not.

$\lambda=0$

| Points, | Planes $t$ | Planes $r$ | Planes $w$ |
|---|---|---|---|
| $t$ | No | Yes | Yes |
| $r$ | Yes | Yes | Yes |
| $w$ | Yes | Yes | No |

$\lambda\neq 0$

| Points, | Planes $t$ | Planes $r$ | Planes $w$ |
|---|---|---|---|
| $t$ | No | Yes | No |
| $r$ | Yes | No | Yes |
| $w$ | No | Yes | No |

If $\lambda=0$, and not otherwise, the point $(w)$ lies on the conic in the trope $(t)$, and is in fact the other point besides $(r)$ where the plane $(r)$ cuts the conic. Also, in this case only, the line joining $(r)$ and $(w)$ is tangent to $ds^2$.

It is characteristic of the case $\lambda=0$ that the point $(r)$ which lies on the conic also lies on the quadric $ds^2$. Since there are four such points, there are four complexes in the co-singular family for which $\lambda=0$.†

4. **Singular lines of the quadratic complex.** If a vector is displaced parallel to itself around an infinitesimal parallelogram it will not in general return to its original position. In this section the cases where it does so return are described in terms of the singular lines of the complex.‡ The Einstein equations and a fundamental tetrahedron of reference are assumed with either $\lambda=0$ or $\lambda\neq 0$. The vector difference between the vector in its final and initial positions is given by

$$\Delta v^i = \gamma_{ij,kl}\omega_l\omega^*{}_k v^j = \tfrac{1}{2}\gamma_{ij,kl}p_{lk}v^j, \tag{4.1}$$

where $\omega_k$ and $\omega_k{}^*$ are the differentials determining the parallelogram. If there is a $v^k$ for which $\Delta v^k$ vanishes for a given $p_{kl}$, then the determinant of

† Jessop, loc. cit., chapter 8.

‡ Jessop, loc. cit., p. 90.

coefficients of $v^k$ is zero, and in fact of rank two. The point $v^k$ can lie anywhere on a certain line $(s_{kl})$. The condition on $(p_{kl})$ turns out to be

$$(4.2) \quad A_{11}^2(p_{01}+p_{23})^2 - \overline{A}_{11}^2(-p_{01}+p_{23})^2 + (\text{cyclic } 1,2,3) = 0.$$

The Plücker coördinates of the line $(s)$ are

$$s_{01} = \gamma_{01,01}p_{23} + \gamma_{01,23}p_{01}, \quad s_{23} = \gamma_{01,01}p_{01} + \gamma_{01,23}p_{23} \qquad (\text{cyclic } 1,2,3).$$

If the line $(p)$ cuts the line $(s)$ then

$$(4.3) \quad A_{11}(p_{01}+p_{23})^2 + \overline{A}_{11}(-p_{01}+p_{23})^2 + (\text{cyclic } 1,2,3) = 0,$$

which is the same as the original complex $\gamma_{ij,kl}p_{ij}p_{kl}=0$. *Equations* (4.2) *and* (4.3) *state that* $(p)$ *is a singular line of the first order.* If in addition $(s)$ belongs to the complex (4.3), then

$$A_{11}^3(p_{01}+p_{23})^2 + \overline{A}_{11}^3(-p_{01}+p_{23})^2 + (\text{cyclic } 1,2,3) = 0,$$

and $(p)$ is a singular line of the second order.

5. Through a point in any $V_4$ there are two-spread elements having the property that if any vector in the element is displaced around a closed path in the element, then the increment of the vector will lie in the element. In the relativity case for empty space there are twenty-six of these. In $P_3$ the Plücker coördinates of the corresponding lines referred to a fundamental tetrahedron are given in the following table:

| | $p_{01}$ | $p_{02}$ | $p_{03}$ | $p_{23}$ | $p_{31}$ | $p_{12}$ |
|---|---|---|---|---|---|---|
| 6 such as | 1, | 0, | 0 | 0, | 0, | 0 |
| 12 " " | 1, | 1, | 0 | 1, | $-1$, | 0 |
| 4 " " | $(A_{22}-A_{33})^{1/2}$, | $e_2(A_{33}-A_{11})^{1/2}$, | $e_3(A_{11}-A_{22})^{1/2}$ | $(A_{22}-A_{33})^{1/2}$, | $e_2(A_{33}-A_{11})^{1/2}$, | $e_3(A_{11}-A_{22})^{1/2}$ |
| 4 " " | $(\overline{A}_{22}-\overline{A}_{33})^{1/2}$, | $e_2(\overline{A}_{33}-\overline{A}_{11})^{1/2}$, | $e_3(\overline{A}_{11}-\overline{A}_{22})^{1/2}$ | $-(\overline{A}_{22}-\overline{A}_{33})^{1/2}$, | $-e_2(\overline{A}_{33}-\overline{A}_{11})^{1/2}$, | $-e_3(\overline{A}_{11}-\overline{A}_{22})^{1/2}$ |

where $e_2$ and $e_3$ are either $+1$ or $-1$. The first eighteen listed above are the edges of the six fundamental tetrahedra which are self-polar with respect to $ds^2$, and the other eight are rulings of $ds^2$. This table was computed from (4.1) and from the equations which state that $v^k$ and $\Delta v^k$ lie on $p_{jk}$.

6. **$\theta$-functions.** This section contains formulas for the components of the curvature in terms of $\theta$-functions of two variables. The coördinates of points on the Kummer surface are expressed uniformly by means of these functions.* The branch points of the corresponding Riemann surface are $A_{11}, A_{22}, A_{33}, -\overline{A}_{11}, -\overline{A}_{22}, -\overline{A}_{33}$, defined in §2, with $A_{jk}=0$ for $j\neq k$. Using $\sum r_k t_k = 0$, the equations (3.4) and (3.5) may be written

* Encyklopädie der Mathematischen Wissenschaften, II B7, p. 747. Humbert, Journal de Mathématiques, (4), vol. 9 (1893), p. 29. Hudson, loc. cit., p. 181. This reference is probably the most convenient in this connection.

$$(6.1)\quad \gamma_{01,01} + \gamma_{02,31} - \gamma_{03,12} = -(\bar{\Omega}_{2,rt}^2 - \Omega_{1,rt}^2 - \Omega_{2,rt}^2 + q_1^2)/Q_1^2,$$

$$(6.2)\quad \gamma_{01,01} - \gamma_{02,31} + \gamma_{03,12} = -(\Omega_{2,rt}^2 - \bar{\Omega}_{1,rt}^2 - \bar{\Omega}_{2,rt}^2 + \bar{q}_1^2)/\bar{Q}_1^2,$$

where

$$\Omega_{1,rt} = r_1 l_0 - r_0 l_1 + r_3 l_2 - r_2 l_3 \qquad \text{(cyclic 1, 2, 3)},$$

$$\bar{\Omega}_{1,rt} = -r_1 l_0 + r_0 l_1 + r_3 l_2 - r_2 l_3 \qquad \text{(cyclic 1, 2, 3)},$$

$$q_1 = r_1 l_0 + r_0 l_1 + r_3 l_2 + r_2 l_3,$$

$$\bar{q}_1 = -r_1 l_0 - r_0 l_1 + r_3 l_2 + r_2 l_3,$$

$$Q_1 = 2(t_0 l_1 + t_2 l_3), \qquad \bar{Q}_1 = 2(-t_0 l_1 + t_2 l_3).$$

The expressions appearing on the right hand sides are the bilinear and quadratic forms discussed in Hudson.* The quadratic forms $Q_1$ and $\bar{Q}_1$ with the eight other similar ones are the squares of $\theta$-functions with zero arguments, with a factor of proportionality $k_1$. Let $u=(u_1, u_2)$ be the parameters of the point $r_k$. Since $(r)$ is on the conic in trope $(t)$, the $\theta$ with zero characteristic vanishes, $\theta_0(u)=0$. Then the bilinear forms in $t_k$ and $r_k$ are the squares of the thetas with argument $(u)$,† with a factor of proportionality $k_2$. Hudson's notation for the characteristics (p. 178) will be used. The zero characteristic is denoted by the symbol $dd$ or 0.

The factor $k_1$ is given by

$$(6.3)\quad 1 = k_1 \theta_{dd}^2(0).$$

If $\lambda \neq 0$, the factor $k_2$ may be found as follows. The sum of the squares of the bilinear forms appearing in any row or column of the table in Hudson (p. 30) is found to be $\sum r^2$ which by (3.7) is equal to $\lambda$. Thus

$$k_2^2 [\theta_{dd}^4(u) + \theta_{ab}^4(u) + \theta_{bc}^4(u) + \theta_{ca}^4(u)] = \lambda.$$

Replacing the bracket by its equivalent in terms of a theta with doubled argument‡

$$(6.4)\quad k_2^2 \theta_{dd}^3(0)\theta_{dd}(2u) = \lambda.$$

Equation (6.1) then becomes

$$(6.5)\quad \gamma_{01,01} + \gamma_{02,31} - \gamma_{03,12} = -k_2^2 [\theta_{ba}^4(u) - \theta_{bc}^4(u) - \theta_{ab}^4(u) + \theta_{ad}^4(u)] / \{k_1^2 \theta_{ad}^4(0)\}.$$

---

* Loc. cit., p. 30.

† Hudson, loc. cit., p. 180.

‡ Schleiermacher, Mathematische Annalen, vol. 50 (1898), p. 200, formula 4.

This bracket also may be simplified by using the thetas with doubled argument. In the Riemann theta formula,* put the characteristics $\rho=0$ and $\sigma=0$ and

$$\begin{aligned} u'+v'+w'+t' &= 2u, & 2u' &= u,\\ u'+v'-w'-t' &= 0, & 2v' &= u,\\ u'-v'+w'-t' &= 0, & 2w' &= u,\\ u'-v'-w'+t' &= 0, & 2t' &= u, \end{aligned}$$

and the bracket becomes $\theta_{ad}^3(0)\theta_{ad}(2u)$. On replacing the $k_1$ and $k_2$ by their values from (6.3) and (6.4), equation (6.5) becomes

$$\gamma_{01,01}+\gamma_{02,31}-\gamma_{03,12}=-\frac{\lambda\theta_{dd}(0)}{\theta_{dd}(2u)}\cdot\frac{\theta_{ad}(2u)}{\theta_{ad}(0)}. \tag{6.6}$$

Thus the desired expressions for the components of the curvature tensor in the case $\lambda\neq 0$ are given in the following table of ten equations; each of the sixteen elements in the first matrix is equal to the corresponding element in the second with the indicated factor:

$$\left\|\begin{array}{llll} -\lambda, & 0, & 0, & 0\\ 0, & \lambda-2\gamma_{01,01}, & \gamma_{03,03}-\gamma_{01,23}+\gamma_{02,31}, & \gamma_{02,02}+\gamma_{03,12}-\gamma_{01,23}\\ 0, & \gamma_{03,03}+\gamma_{01,23}-\gamma_{02,31}, & \lambda-2\gamma_{02,02}, & \gamma_{01,01}-\gamma_{02,31}+\gamma_{03,12}\\ 0, & \gamma_{02,02}-\gamma_{03,12}+\gamma_{01,23}, & \gamma_{01,01}+\gamma_{02,31}-\gamma_{03,12}, & \lambda-2\gamma_{03,03} \end{array}\right\|$$

$$=\frac{-\lambda}{\binom{00}{00}}\left\|\begin{array}{llll} \binom{00}{00}, & 0, & 0, & 0\\ 0, & \binom{00}{01}, & -\binom{01}{10}, & \binom{11}{00}\\ 0, & \binom{01}{00}, & \binom{00}{11}, & -\binom{10}{01}\\ 0, & -\binom{11}{11}, & \binom{10}{00}, & \binom{00}{10} \end{array}\right\|$$

with

$$\gamma_{ij,ik}=0 \quad (j\neq k); \qquad \gamma_{01,01}=\gamma_{23,23} \quad \text{(cyclic 1, 2, 3)}$$

where the symbol $\binom{\alpha\beta}{\gamma\delta}$ denotes the quotient

$$\theta\left[\begin{smallmatrix}\alpha\beta\\ \gamma\delta\end{smallmatrix}\right](2u)/\theta\left[\begin{smallmatrix}\alpha\beta\\ \gamma\delta\end{smallmatrix}\right](0).$$

The derivation of the other equations in the table is similar to that of (6.6).

If $\lambda=0$, the point $(r)$ whose parameters are $(u)$ lies on $ds^2$, and therefore $\theta_{dd}(2u)=0$.† The same equations are true in this case, with a different fac-

* Krazer, *Theorie der Zweifach Unendlichen Thetareihen*, pp. 2, 9 and 65 (the case 8−10−14+2). See also the Encyclopedia article above, p. 673.

† Hudson, loc. cit., p. 189.

tor of proportionality. This factor cannot be found in terms of the $\gamma_{ij,kl}$ at a single point in $V_4$.

7. **Reality relations.** In the formula

$$ds^2 = \omega_0^2 + \omega_1^2 + \omega_2^2 + \omega_3^2,$$

$\omega_0$ is a pure imaginary and the other $\omega_k$ are real, for real values of the $x^\alpha$ and $dx^\alpha$. This convention of signs, while not the usual one, is used so that the writers on the line complex may be more closely followed. The coefficients of rotation $\gamma_{ijk}$ and the curvatures $\gamma_{ij,kl}$ are real if the index 0 occurs an even number of times, and are otherwise pure imaginary. In the preceding sections the bar over a letter has meant a different function. In the physical case the bar is to mean conjugate imaginary.*

A point in $P_3$, or vector in $V_4$, is real if the ratio of the $j$th coördinate ($j=1, 2, 3$) to the 0th is pure imaginary. From (3.1) and (3.2) it can be shown that each of the four points $(r)$ in the four real tropes is real. These are the points on the main diagonal of the table before (3.4). Only one of the six fundamental tetrahedra self-polar with respect to $ds^2$ has four real vertices.

If the $\theta$-functions are defined as in Hudson (p. 176) and in Baker's book on Abelian functions (p. 248),

$$\theta \begin{bmatrix} \alpha_1, \alpha_2 \\ \beta_1, \beta_2 \end{bmatrix} (u) = \sum \exp \left\{ 2\pi i (n + \tfrac{1}{2}\alpha)\ (u + \tfrac{1}{2}\beta) + \pi i \tau (n + \tfrac{1}{2}\alpha)^2 \right\},$$

where $\alpha$, $\beta$, $n$ and $u$ are row letters and $\tau$ is a two-by-two matrix, then $\tau_{11}$ is the negative conjugate of $\tau_{22}$, and $\tau_{12}$ is pure imaginary.

This is now proved by considering the hyperelliptic integrals of the first kind:

$$w_1 = \int \frac{dz}{R}, \qquad w_2 = \int \frac{z\,dz}{R},$$

$$R = [(A_{11} - z)(-\overline{A}_{11} - z)(A_{22} - z)(-\overline{A}_{22} - z)(A_{33} - z)(-\overline{A}_{33} - z)]^{1/2}.$$

The six branch points are situated symmetrically with respect to the axis of imaginaries. This follows from the expressions for the $A_{jj}$ and $\overline{A}_{jj}$ in §2, and from the reality criterion near the beginning of this section. In constructing the Riemann surface take, for example, the $a_1$ and $b_1$ cuts† wholly in the first and fourth quadrants. Then take the $a_2$ and $b_2$ cuts symmetrically in the second and third. The moduli of periodicity are seen to have the reality relations given by the following table:

---

* The situation is discussed in Hudson's book on p. 126, under the case IVa.

† See Goursat-Appell, *Théorie des Fonctions Algébriques et de leurs Intégrales*, Chapter III, pp. 132, 133, 151 for notation and details.

| | $a_1$ | $a_2$ | $b_1$ | $b_2$ |
|---|---|---|---|---|
| $w_1$ | $C_1$ | $-\overline{C}_1$ | $D_1$ | $+\overline{D}_1$ |
| $w_2$ | $C_2$ | $+\overline{C}_2$ | $D_2$ | $-\overline{D}_2$ |

When the integrals are normalized in the usual way the required result is obtained. From this result and from the definition of the thetas it follows that $\theta_{dd}(0)$ is real, and then that the ten non-vanishing thetas with zero argument, when arranged as in §6, form an Hermitian matrix.

8. **The Schwarzschild solution.** This does not come under the preceding discussion because the complex is not of type [111111]. In fact if $x_0$ is the time, and $x_1$ is taken radially from the central mass, then

$$A_{11} = \overline{A}_{11} = -\frac{2m}{x_1^3}, \quad A_{22} = \overline{A}_{22} = A_{33} = \overline{A}_{33} = \frac{m}{x_1^3}.$$

The complex is of type [(11)(11)11] and is a special case of the harmonic complex.*

The detailed examination of the complex (2.7) shows that the field equations exclude the cases of elementary divisors listed below. All except these twelve are algebraically possible:

| | | | | |
|---|---|---|---|---|
| [(222)], | [(411)], | [(42)], | [(51)], | [6], |
| [(22)2], | [4(11)], | [42], | [51], | |
| [222], | [(41)1], | | | |
| | [411]. | | | |

* Jessop, loc. cit., pp. 209, and 111.

LEHIGH UNIVERSITY,
BETHLEHEM, PA.

# GENERALIZED DIFFERENTIATION*

BY

EMIL L. POST

## INTRODUCTION

The history of the subject of generalized differentiation can be traced back to Leibnitz.† In the earlier literature the term fractional differentiation is used as an alternative, and the aim was to generalize the concept of the $n$th derivative of a function to non-integral values of $n$. Four different attacks on this problem may be noted. The earliest was that of Liouville, who expanded the functions operated upon in series of exponentials, and assumed, as a basis, $D^n e^{ax} = a^n e^{ax}$, where $D$ symbolizes differentiation. Riemann considered power series with non-integral exponents as analogues of Taylor's series, and through their coefficients was led to the expression of generalized derivatives in terms of a definite integral plus an infinite series with arbitrary constant coefficients.‡ Liouville's and Riemann's results proved to be in disagreement. Prior to the publication of Riemann's work, Grünwald was led by the restrictions of Liouville's method to generalize directly the definition of a derivative as the limit of a finite difference quotient, and, by rigorous methods, also arrived at definite integral formulas.§ Grünwald's definition involved the idea of differentiation between limits which later resulted in the coördination of Liouville's and Riemann's results. This, for example, was effected by Krug,¶ who introduced a new development, based on Cauchy's contour integral for ordinary derivatives, which also involved limits of differentiation, in terms of which he showed that Riemann's definite integral corresponded to finite lower limit, Liouville's development to lower limit $-\infty$. The Riemann-Grünwald definite integral form has become standard in the literature, and has been intensively studied.‖

---

* Presented to the Society, October 27, 1923; received by the editors June 29, 1929.

† For references, see S. Pincherle, *Equations et opérations fonctionnelles*, Encyclopédie des Sciences Mathématiques, Paris, 1912, tome 2, vol. 5, fasc. 1, pp. 1–81; also Eugene Stephens, *Symbolic calculus. Bibliography on general (or fractional) differentiation*, Washington University Studies, vol. 12 (1925), No. 2, pp. 137–152.

‡ B. Riemann, Gesammelte Mathematische Werke, Leipzig, 1876, pp. 331–344.

§ K. A. Grünwald, Zeitschrift für Mathematik und Physik, vol. 12 (1867), pp. 441–480.

¶ A. Krug, $d^n f(x)/dx^n$ *regarded as a function of* $n$, Akademie der Wissenschaften, Wien, Denkschriften, Mathematisch-Naturwissenschaftliche Klasse, vol. 57 (1890), pp. 151–228.

‖ E.g. by A. Marchaud, *Sur les dérivées et sur les différences des fonctions de variables réelles*, Journal de Mathématiques, (9), vol. 6 (1927), pp. 337–425.

Another trend was introduced by the symbolic treatment of linear differential equations with constant coefficients as exemplified by the work of Boole. The polynomial operators thus occurring led naturally to the concept of an arbitrary operator $f(D)$, which was to be formally expanded in powers of $D$, and thus applied to the operand. This treatment has since been established on a rigorous basis for operators $f(D)$ corresponding to entire transcendental functions $f(z)$ of genus zero, operating upon functions which are analytic in a given region.* Boole's methods for linear differential equations with variable coefficients were extended by other writers to yield formal solutions in terms of operators algebraic in $D$; but in many cases no further attempt was made to assign a meaning to such expressions.

We may think of the more recent work of Heaviside as the next step in this development.† Of course Heaviside's contribution assumes an importance far beyond this formal juggling of symbols, through its application to important physical problems, and its skillful methods for evaluating the operations that are used. This "operational calculus," however, was developed with physical intuition, rather than mathematical rigor, as guide. A more rigorous mathematical basis has since been supplied by Carson in terms of solutions of Laplace integral equations, and their use in a definite integral formula.‡ Carson's formulas include the Riemann-Grünwald definite integral for $D^n$ as a special case when the real part of $n$ is less than one; but his treatment is quite unrelated to the theory of entire operators of genus zero.§

In the present paper Grünwald's method of arriving at a definition for operators $D^n$ is carried forward by means of an artifice of Arbogast to yield a definition of generalized differentiation for operators $f(D)$.¶ This definition is shown to include Carson's operators, and entire operators of genus zero, as special cases.‖ The major part of the paper is devoted to operators $f(D)$,

* C. Bourlet, *Sur les opérations en général et les équations linéaires différentielles d'ordre infini*, Annales de l'Ecole Normale, (3), vol. 33 (1897), pp. 133–190.

J. F. Ritt, *On a general class of linear homogeneous differential equations of infinite order with constant coefficients*, these Transactions, vol. 18 (1917), pp. 27–49.

† Oliver Heaviside, *Electromagnetic Theory*, London, 1922, vol. 2, chapters 7, 8.

‡ J. R. Carson, *The Heaviside operational calculus*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 43–68; also numerous papers in the Bell System Technical Journal.

§ Another theory is developed by Norbert Wiener: *The operational calculus*, Mathematische Annalen, vol. 95 (1926), pp. 557–584.

¶ Our definition may therefore be called the extended Grünwald definition.

‖ This is not strictly correct as far as Carson's development is concerned, since we derive his formulas only under certain hypotheses on the functions involved. It should be noted, however, that Carson does not explicitly state the domain of applicability of these formulas, and that the hypotheses in question are of considerable generality.

termed of type zero, which correspond to functions of a complex variable, $f(z)$, which are analytic in a certain sector of the $z$-plane of angle greater than $\pi$, and whose moduli satisfy within this sector the Poincaré inequality for entire functions of genus zero.* Existence theorems are established for such operators, and certain formal properties are investigated, such as the law of successive operations, and the generalized Leibnitz theorem for differentiation of a product.† The last three sections are chiefly devoted to operators given by a Laplace integral. For these we have only established the existence theorem, and investigated the application to Carson's development where not $f(z)$, but $f(z)/z$, is given by a Laplace integral.

Next to the definition of generalized differentiation itself, the writer wishes to call attention to the associated operator $A[f](t)$, defined by

$$A[f](t) = \lim_{\substack{\Delta x \to +0 \\ r\Delta x = t}} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r!\Delta x^{r+1}}.$$

The existence of this limit, for the cases considered, serves as the basis of our theory. It possesses the notable property of inverting the Laplace transformation, and enjoys numerous formal relations which flow directly from its definition. For $f(D)$ of type zero, it can be expressed by a contour integral, which is the Fourier integral with the vertical line contour replaced by two half-lines running into the negative half of the $z$-plane. The exponential factor of the integrand thus acquires a potency for convergence which should be of considerable use in most practical applications, where the loss of generality incurred by requiring analyticity in a sector of angle greater than $\pi$, as opposed to analyticity in a half-plane, is irrelevant, due to the presence of but a finite number of singularities.

In order to bring this paper to a conclusion, many topics have been excluded either because they involve an extension of our definition, or because they depend on the formulas flowing from our definition, rather than on the definition itself. Among these omitted topics are an extension of our definition to complex limits of differentiation, the application to Volterra's functions of composition of the closed cycle group, and the derivation of the Heaviside expansions.

* By thus breaking away from the classic assumption of analyticity in a half-plane, the restriction on the modulus of $f(z)$ is so greatly lightened (see footnote following proof of Theorem I) that we are able to include all operators $f(D)$ for which $f(z)$ is an entire function of genus zero, as well as all operators for which $f(z)$ is algebraic. The half-plane assumption would exclude these classes, as such, and would also hinder the theory by not permitting indiscriminate differentiation with respect to the upper limit.

† This treatment of the extended Grünwald definition follows very closely Grünwald's development of the original definition.

1. **The definition.** Our definition of generalized differentiation grows out of the expression of the $n$th derivative of a function as the limit of its $n$th difference quotient. For our purpose, this expression is best written in the form

$$D^n\phi(x) = \lim_{\Delta x \to 0} \frac{\phi(x) - n\phi(x-\Delta x) + [n(n-1)/2!]\phi(x-2\Delta x) - \cdots + (-1)^n\phi(x-n\Delta x)}{\Delta x^n},$$

where $\Delta x$ is the negative of the increment of $x$ as ordinarily defined. Since the right hand member of this equation continues to have meaning when $n$ is not a positive integer, it can be used to define $D^n\phi(x)$ for arbitrary $n$.

When $n=-1$, the suggested definition becomes

$$D^{-1}\phi(x) = \lim_{\Delta x \to +0} [\phi(x)\Delta x + \phi(x-\Delta x)\Delta x + \phi(x-2\Delta x)\Delta x + \cdots],$$

i.e., the limit of an infinite series. If, however, we arbitrarily terminate the series at the $(p+1)$st term, where $x_0 < x - p\Delta x \leqq x_0 + \Delta x$, the result, at least for $\phi(x)$ continuous, will be the definite integral of $\phi(x)$ with finite lower limit $x_0$. Replacing the upper limit by $X$, we shall use the notation $\{D^{-1}\}_{x_0}^{X}\phi(x)$ for the limit of the finite sum, $\{D^{-1}\}[_{-\infty}^{X}\phi(x)$ for the limit of the infinite series. Similarly for $D^n$, $n$ arbitrary.*

To extend this definition still further, let us momentarily introduce the operator $E^m$, defined by the relation $E^m\phi(x) = \phi(x+m)$. The reader will have seen the analogy between the above expression for $D^n\phi(x)$ and the binomial series expansion. Through the operator $E^m$, this analogy becomes formal, and we can write, apart from refinements,

$$D^n\phi(x) = \lim_{\Delta x \to 0} \left(\frac{1 - E^{-\Delta x}}{\Delta x}\right)^n \phi(x) = \lim_{\Delta x \to 0} \left(\frac{\Delta}{\Delta x}\right)^n \phi(x).$$

This immediately suggests the desired definition of $f(D)\phi(x)$, i.e.,

$$f(D)\phi(x) = \lim_{\Delta x \to 0} f\left(\frac{\Delta}{\Delta x}\right)\phi(x) = \lim_{\Delta x \to 0} f\left(\frac{1 - E^{-\Delta x}}{\Delta x}\right)\phi(x),$$

where $f(1/\Delta x - E^{-\Delta x}/\Delta x)$ is to be expanded by Taylor's series, and formally applied to $\phi(x)$. As in the case of $D^{-1}$ the limit of the infinite series will be written $\{f(D)\}_{-\infty}^{X}\phi(x)$. In the case of finite lower limit $x_0$, the choice of $p$ through the inequalities $x_0 < X - p\Delta x \leqq x_0 + \Delta x$ allows $\Delta x$ to approach zero independently of $X - x_0$, except for sign. $p$ is then the largest integer less

* Up to this point the writer essentially retraces Grünwald's argument.

than $(X-x_0)/\Delta x$. We shall use the notation $p=\{(X-x_0)/\Delta x\}$. Though otherwise arbitrary, $\Delta x$ must have the same sign as $X-x_0$. In this paper $\Delta x$ will approach zero positively, that is, $X$ will be greater than $x_0$. Our completed definitions thus become

$$(1)\quad \{f(D)\}_{x_0}^{X}\phi(x) = \lim_{\Delta x\to+0}\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x) = \lim_{\Delta x\to+0}\left\{f\left(\frac{1-E^{-\Delta x}}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x)$$

$$= \lim_{\substack{\Delta x\to+0\\ p=\{(X-x_0)/\Delta x\}}}\left[f\left(\frac{1}{\Delta x}\right)\phi(X) - \frac{f'\left(\frac{1}{\Delta x}\right)}{1!\Delta x}\phi(X-\Delta x) + \cdots + \frac{(-1)^p f^{(p)}\left(\frac{1}{\Delta x}\right)}{p!\Delta x^p}\phi(X-p\Delta x)\right],$$

$$(2)\quad \{f(D)\}_{-\infty}^{X}\phi(x) = \lim_{\Delta x\to+0}\left[f\left(\frac{1}{\Delta x}\right)\phi(X) - \frac{f'\left(\frac{1}{\Delta x}\right)}{1!\Delta x}\phi(X-\Delta x) + \cdots\right].$$

The existence of the limits involved in these definitions is proved for various classes of operators in subsequent sections of the paper. In certain simple cases they can be evaluated directly. For example, consider $f(D)=\log D$, $\phi(x)\equiv 1$, and $x_0$ finite. Then by (1)

$$\left\{\log\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}1 = \log\left(\frac{1}{\Delta x}\right) - 1 - \frac{1}{2} - \cdots - \frac{1}{p}$$

$$= -\left[1+\frac{1}{2}+\cdots\frac{1}{p}-\log p\right] - \log(p\Delta x).$$

As $\Delta x\to+0$, $p$ increases indefinitely, while $p\Delta x\to X-x_0$. Furthermore, as $p\to\infty$, the bracket has for limit $\gamma$, the Eulerian constant. We thus find

$$(3)\qquad \{\log(D)\}_{x_0}^{X}1 = -\gamma-\log(X-x_0).$$

Besides proving existence theorems, with their associated formulas, we shall be concerned with the verification of the formal laws of generalized differentiation for our definitions. Of these the most important is the law of successive operations. This can be immediately verified for the finite difference operators on which our generalized derivatives are based. Assuming the necessary number of derivatives of $f_1$ and $f_2$, we have

$$\left\{f_1\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\left\{f_2\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{x}\phi(\xi) = f_1\left(\frac{1}{\Delta x}\right)\left[f_2\left(\frac{1}{\Delta x}\right)\phi(X)\right.$$

$$\left. - \frac{f_2'\left(\frac{1}{\Delta x}\right)}{1!\Delta x}\phi(X-\Delta x) + \cdots + \frac{(-1)^p f_2^{(p)}\left(\frac{1}{\Delta x}\right)}{p!\Delta x^p}\phi(X-p\Delta x)\right]$$

$$- \frac{f_1'\left(\frac{1}{\Delta x}\right)}{1!\Delta x}\left[f_2\left(\frac{1}{\Delta x}\right)\phi(X-\Delta x) - \cdots + \frac{(-1)^{p-1} f_2^{(p-1)}\left(\frac{1}{\Delta x}\right)}{(p-1)!\Delta x^{p-1}}\phi(X-p\Delta x)\right]$$

$$+ \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

$$+ \frac{(-1)^p f_1^{(p)}\left(\frac{1}{\Delta x}\right)}{p!\Delta x^p}\left[f_2\left(\frac{1}{\Delta x}\right)\phi(X-p\Delta x)\right]$$

$$= \left[f_1\left(\frac{1}{\Delta x}\right)f_2\left(\frac{1}{\Delta x}\right)\right]\phi(X)$$

$$- \frac{\left[f_1\left(\frac{1}{\Delta x}\right)f_2'\left(\frac{1}{\Delta x}\right) + f_1'\left(\frac{1}{\Delta x}\right)f_2\left(\frac{1}{\Delta x}\right)\right]}{1!\Delta x}\phi(X-\Delta x)$$

$$+ \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .$$

$$+ (-1)^p \frac{\left[f_1\left(\frac{1}{\Delta x}\right)f_2^{(p)}\left(\frac{1}{\Delta x}\right) + \frac{p}{1!}f_1'\left(\frac{1}{\Delta x}\right)f_2^{(p-1)}\left(\frac{1}{\Delta x}\right) + \cdots + f_1^{(p)}\left(\frac{1}{\Delta x}\right)f_2\left(\frac{1}{\Delta x}\right)\right]}{p!\Delta x^p}$$

$$\cdot \phi(X-p\Delta x),$$

so that, by Leibnitz's theorem and our definition, we obtain

$$\left\{f_1\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\left\{f_2\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{x}\phi(\xi) = \left\{f_1\left(\frac{\Delta}{\Delta x}\right)f_2\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x). \tag{4}$$

The character of definition (1) is more apparent in the following form:

$$\{f(D)\}_{x_0}^{X}\phi(x) = \lim_{\substack{\Delta x \to +0 \\ p=\{(X-x_0)/\Delta x\}}} \sum_{r=0}^{p} A[f](r, \Delta x)\phi(X - r\Delta x)\Delta x;$$

$$A[f](r, \Delta x) = \frac{(-1)^r f^{(r)}\left(\frac{1}{\Delta x}\right)}{r!\Delta x^{r+1}}. \tag{5}$$

(5) would lead directly to a definite integral if in place of $A[f](r, \Delta x)$ we had a function of $r\Delta x$. The way in which this difficulty can be partly overcome may be indicated by means of the operator $D^n$, $n$ arbitrary. Using the notation $A[f(u)](r, \Delta x)$ when the form of $f$ is specified, we have, in this case,

$$A[u^n](r, \Delta x) = \frac{(-1)^r n(n-1)\cdots(n-r+1)}{r!\Delta x^{n+1}}$$

$$= \frac{(-n)(-n+1)\cdots(-n+r-1)}{r!r^{-n-1}}(r\Delta x)^{-n-1}.$$

The Gaussian form of the gamma function gives

$$\lim_{r\to\infty} \frac{(-n)(-n+1)\cdots(-n+r-1)}{r!r^{-n-1}} = \frac{1}{\Gamma(-n)}.$$

If at the same time $\Delta x \to +0$ in such a way that $r\Delta x \to t$, $t>0$, it will follow that $A[u^n](r, \Delta x)$ will approach a function of $t$ as limit. Symbolizing this function by $A[u^n](t)$, we thus find

$$A[u^n](t) = \lim_{\substack{\Delta x\to+0 \\ r\Delta x\to t}} A[u^n](r, \Delta x) = \frac{t^{-n-1}}{\Gamma(-n)}. \tag{6}$$

More generally we shall look for the existence of a limit function $A[f](t)$ corresponding to $A[f](r, \Delta x)$. When such a limit exists, then for $\Delta x$ sufficiently small, and $r\Delta x$ greater than some positive $\epsilon$, $A[f](r, \Delta x)$ will be approximated by $A[f](r\Delta x)$; and so, (5) will lead in part to a definite integral. Nevertheless, the restriction $r\Delta x > \epsilon$ leaves terms of (5) with small $r$ unaccounted for. These will usually require separate treatment, where, however, the $A[f](t)$'s of operators connected with $f(D)$ will be used. The first step in our theory is therefore the establishing of the existence of $A[f](t)$ for a sufficiently wide class of operators $f(D)$. We pause a moment to consider a certain limit criterion which will unify the last stages of many of our proofs.

2. **Limit criterion.** *If $F(\Delta x)$ can be expressed as a sum $G(\Delta x, \nu)+H(\Delta x, \nu)$, $\nu$ independent of $\Delta x$, such that*

$$\lim_{\Delta x\to+0} G(\Delta x, \nu) = g(\nu), \quad \limsup_{\Delta x\to+0} |H(\Delta x, \nu)| \leqq h(\nu), \quad \lim_{\nu\to\nu_0} h(\nu) = 0,$$

*then $\lim_{\Delta x\to+0} F(\Delta x)$ exists, and is given by*

$$\lim_{\Delta x\to+0} F(\Delta x) = \lim_{\nu\to\nu_0} g(\nu). \tag{7}$$

In fact the stated conditions show that

$$\limsup_{\Delta x \to +0} F(\Delta x) \leqq g(\nu) + h(\nu), \quad \liminf_{\Delta x \to +0} F(\Delta x) \geqq g(\nu) - h(\nu).$$

As these upper and lower limits are independent of $\nu$, we find, by letting $\nu \to \nu_0$, and observing that $h(\nu) \to 0$,

$$\limsup_{\Delta x \to +0} F(\Delta x) = \liminf_{\Delta x \to +0} F(\Delta x) = \lim_{\nu \to \nu_0} g(\nu).$$

Since for any one $\nu$ the values of $g(\nu)+h(\nu)$ and $g(\nu)-h(\nu)$ are finite, the above inequalities yield the

COROLLARY. *Under the given hypothesis* $\lim_{\Delta x \to +0} F(\Delta x)$, *and hence also* $\lim_{\nu \to \nu_0} g(\nu)$, *is finite.**

3. **Existence of $A[f](t)$ for $f(D)$ of type zero.** An operator $f(D)$, corresponding to a function of a complex variable $f(z)$, will be said to be of *type zero* when $f(z)$ satisfies the following two conditions:

(a) *$f(z)$ is analytic in a sector of the z-plane of angle greater than $\pi$ bisected by the positive direction of the axis of reals,*

(b) *for each real and positive $\kappa$, however small, there corresponds a real and positive $K$, such that*

$$|f(z)| \leqq Ke^{\kappa|z|},$$

*for every $z$ in the analytic sector.*

We shall designate the positive acute angle made by the sides of the sector with the negative direction of the axis of reals by $\alpha$.

Operators of finite order, defined later, which include all operators whose $f(z)$ is algebraic, are also of type zero. Other examples are $e^{aD^m}$, $m<1$, and operators for which $f(z)$ is an entire function of genus zero.

We shall now prove the fundamental

THEOREM I. *If $f(D)$ is of type zero, $A[f](r, \Delta x)$ approaches a finite limit as $\Delta x \to +0$, $r\Delta x \to t$, for every real and positive $t$. The resulting function of $t$, $A[f](t)$, is given by the formula*

$$A[f](t) = \frac{1}{2\pi i}\int_C e^{tz} f(z)dz, \tag{8}$$

*where $C$ is formed by two rays within the analytic sector and parallel to its edges, with common end point on the axis of reals, and traversed so that, along it, the imaginary part of $z$ increases.*

The proof is based on Cauchy's second integral formula. For $\Delta x$ sufficiently small, $1/\Delta x$ will be in the analytic sector posited by condition (a).

* Henceforth, existence will include finiteness.

If $C'''$ is a simple contour about $1/\Delta x$, traversed counterclockwise, and also contained in this sector, then, by the formula in question, we shall have

$$f^{(r)}\left(\frac{1}{\Delta x}\right) = \frac{r!}{2\pi i}\int_{C'''} \frac{f(z)dz}{\left(z - \frac{1}{\Delta x}\right)^{r+1}};$$

and so, by the definition of $A[f](r, \Delta x)$ given in (5),

$$A[f](r, \Delta x) = -\frac{1}{2\pi i}\int_{C'''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}}.$$

Assuming $\Delta x$ sufficiently small, we can choose for $C'''$ the contour formed by $C''$, an arc of a circle, center the origin, radius $k/\Delta x$, $k>1$, joined to $C'$, the finite portion of $C$ cut off by this circle. If $C''$ is traversed in the same direction as $C'''$, but $C'$ in the opposite direction, i.e., in the same direction as $C$, we can write schematically

$$A[f](r, \Delta x) = -\frac{1}{2\pi i}\int_{C''} + \frac{1}{2\pi i}\int_{C'}.$$

Along $C''$ we have $|z| = k/\Delta x$. Hence $|1 - z\Delta x| \geqq k - 1$, so that, by inequality (b), we find

$$\left|\int_{C''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}}\right| < \frac{K \cdot 2\pi k}{\Delta x(k-1)}\left[\frac{e^{k\kappa/(r\Delta x)}}{k-1}\right]^r.$$

Choose $k>2$, and then $\kappa$ so that $e^{k\kappa/t}/(k-1)<1$. The right hand member of the inequality will then approach zero as limit as $\Delta x \to +0, r\Delta x \to t$. The $C''$ contribution can therefore be neglected.

Now break up $C'$ into $C_{l,m}$ and $C'_{l,m}$, where $C_{l,m}$ extends distances $l$ and $m$ respectively along the two segments of $C'$ from their meeting point, while $C'_{l,m}$ consists of the rest of $C'$. For $l$ and $m$ sufficiently large, $R(z)$, the real part of $z$, will be negative along $C'_{l,m}$, and we shall have

$$\left|\int_{C'_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}}\right| < K\int_{C'_{l,m}} \frac{e^{\kappa|z|}\,|dz|}{[1 - R(z)\Delta x]^{r+1}}.$$

If we set

$$1 - R(z)\Delta x = e^{-\lambda R(z)\Delta x},$$

we observe that $\lambda$ stays positive and decreases monotonically as $-R(z)\Delta x$ increases. Now the largest value of $-R(z)\Delta x$ along any one $C'_{l,m}$ corresponds to $|z| = k/\Delta x$. As $\Delta x \to +0$ this largest value approaches $k \cos \alpha$ as limit. Hence for all $\Delta x$'s sufficiently small, and all corresponding $C'_{l,m}$'s, $-R(z)\Delta x$

remains less than some fixed positive quantity. The corresponding $\lambda$'s therefore have a positive lower bound $\lambda_0$. Since $\lambda \geqq \lambda_0 > 0$, we shall thereby have

$$1 - R(z)\Delta x > e^{-\lambda_0 R(z)\Delta x},$$

and so

$$\left| \int_{C'_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < K \int_{C'_{l,m}} e^{\kappa|z| + \lambda_0 (r+1)\Delta x \cdot R(z)} \mid dz \mid .$$

Since $R(z)/|z| \to -\cos\alpha$, as $|z| \to \infty$, if we choose $\kappa$ less than $\lambda_0 l \cos\alpha$ the last integral will be less than

$$\int_{C'_{l,m}} e^{-\mu|z|} \mid dz \mid ,$$

with fixed positive $\mu$, for $l$ and $m$ sufficiently large, and $(r+1)\Delta x$ sufficiently near $t$. The contour $C'_{l,m}$ depends on $\Delta x$. If we replace it by $\overline{C}'_{l,m}$, which consists of $C$ with $C_{l,m}$ removed, and of which $C'_{l,m}$ is a part, we observe that the resulting integral converges, and so approaches zero as limit as $l$ and $m$ increase indefinitely. We therefore have

$$\limsup_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \left| \int_{C'_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| \leqq K \int_{\overline{C}'_{l,m}} e^{-\mu|z|} \mid dz \mid , \quad \lim_{\substack{l \to \infty \\ m \to \infty}} \int_{\overline{C}'_{l,m}} e^{-\mu|z|} \mid dz \mid = 0.$$

Finally $f(z)/(1-z\Delta x)^{r+1}$ uniformly approaches $e^{tz}f(z)$ along $C_{l,m}$. Hence

$$\lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \int_{C_{l,m}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} = \int_{C_{l,m}} e^{tz} f(z)dz.$$

We can therefore apply our limit criterion,* and obtain

$$\lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \int_{C'} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} = \lim_{\substack{l \to \infty \\ m \to \infty}} \int_{C_{l,m}} e^{tz} f(z)dz = \int_{C} e^{tz} f(z)dz.$$

Since we saw that the integral along $C''$ could be neglected, this establishes both the existence of $A[f](t)$, and the formula given for it.†

* This was stated for one independent variable $\Delta x$. It can obviously be extended to any number of independent variables, in this case two.

† If the analyticity condition imposed on $f(z)$ be weakened to analyticity in the half-plane to the right of some line $R(z)=a$, while within that half-plane the modulus of $f(z)$ satisfies the far stronger inequality

$$|f(x)| \leqq K/|z|^{1+\epsilon}, \epsilon > 0,$$

then essentially the same proof will yield the existence of $A[f](t)$, and its expression by means of a Fourier integral. Furthermore, to anticipate the later developments, the argument leading to (48) in §11 can be duplicated in this case, so that our methods would yield the classic solution of the Laplace integral equation.

By real-variable methods, it can be proved that if $A[f](t)$ exists for each $t$ in a closed interval $(t_1, t_2)$, then, without any further hypothesis on the function $f$, or even on the form of $A[f](r, \Delta x)$, the following consequences hold:

(a) $A[f](t)$ is continuous in the closed interval $(t_1, t_2)$;

(b) for each positive $\epsilon$ however small, there corresponds a positive $\eta$, such that, for $\Delta x < \eta$,

$$| A[f](r, \Delta x) - A[f](r\Delta x) | < \epsilon$$

for all $r$'s for which $r\Delta x$ is in the interval $(t_1, t_2)$.*

From our definition we have, for $x_0 < x_1 < X$,

$$\left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_0}^{X} \phi(x) = \left\{ f\left(\frac{\Delta}{\Delta x}\right) \right\}_{x_1}^{X} \phi(x) + \sum_{r=q+1}^{p} A[f](r, \Delta x)\phi(X - r\Delta x)\Delta x,$$

where $p = \{(X-x_0)/\Delta x\}$, $q = \{(X-x_1)/\Delta x\}$. If we set $X - r\Delta x = x$, then, as $r$ varies from $q+1$ to $p$, $x$ stays in the closed interval $(x_0, x_1)$, while $r\Delta x$ stays in the interval $(X-x_1, X-x_0)$. Suppose then that $\phi(x)$ is continuous in $(x_0, x_1)$, while $A[f](t)$ exists in $(X-x_1, X-x_0)$. Then by (b), $A[f](r, \Delta x)$ can be replaced by $A[f](r\Delta x)$, i.e., by $A[f](X-x)$, in the above sum. But by (a), $A[f](X-x)$ will be continuous in $x$ in the interval $(x_0, x_1)$. The new sum will therefore lead to a definite integral as $\Delta x \to +0$. Hence

THEOREM II. *If $A[f](t)$ exists in the interval $(X-x_1, X-x_0)$, where $x_0 < x_1 < X$ and if $\phi(x)$ is continuous in the interval $(x_0, x_1)$, then*

$$\{f(D)\}_{x_0}^{X} \phi(x) = \{f(D)\}_{x_1}^{X} \phi(x) + \int_{x_0}^{x_1} A[f](X - x)\phi(x)dx, \tag{9}$$

*provided either of the indicated operations exist.*

When $f(D)$ is of type zero $A[f](t)$ exists for all positive $t$'s. If then $\phi(x)$ is continuous in $(x_0, X)$, (9) will be valid with $x_1$ anywhere in this interval.

* The proof runs as follows: Choose any positive $\epsilon$. For each point $t'$ of the interval $(t_1, t_2)$, $A[f](t')$ exists, and is defined as $\lim A[f](r, \Delta x)$ as $\Delta x \to +0$, $r\Delta x \to t'$. Hence for each $t'$ of $(t_1, t_2)$ there is a positive $\eta'$ such that $|A[f](r, \Delta x) - A[f](t')| < \epsilon/2$ provided $\Delta x < \eta'$ and $|r\Delta x - t'| < \eta'$. By letting $\Delta x \to +0$ and $r\Delta x \to t$ we obtain $|A[f](t) - A[f](t')| \leqq \epsilon/2$ provided $|t - t'| < \eta'$ and $t$ is in $(t_1, t_2)$. Hence $A[f](t)$ is continuous at every point $t'$ of $(t_1, t_2)$.

Consider now the open intervals $(t'-\eta', t'+\eta')$ thus associated with the above pairs $(t', \eta')$. Every point of $(t_1, t_2)$ is in fact the midpoint of such an interval. Hence by the Heine-Borel theorem a finite number of these intervals suffice to cover $(t_1, t_2)$. Each interval uniquely determines the corresponding $t'$ and $\eta'$. Let $\eta$ be the smallest of the $\eta'$'s of this finite set of intervals, and let $\Delta x < \eta$. Any $r\Delta x$ in $(t_1, t_2)$ will be in one of these intervals. Since, for the $(t', \eta')$ of this interval, $\Delta x < \eta \leqq \eta'$ and $|r\Delta x - t'| < \eta'$, we will have $|A[f](r, \Delta x) - A[f](t')| < \epsilon/2$ and $|A[f](r\Delta x) - A[f](t')| \leqq \epsilon/2$. Hence $|A[f](r, \Delta x) - A[f](r\Delta x)| < \epsilon$. That is, this inequality holds for every $r$ and $\Delta x$ for which $\Delta x < \eta$ and $r\Delta x$ is in $(t_1, t_2)$.

The suggestion made at the end of §1 is thus verified. Before we see how to deal with $\{f(D)\}_x^X\phi(x)$, we shall consider $\phi(x)$ a polynomial. In this case, $\{f(D)\}_{x_0}^X\phi(x)$ can be immediately evaluated.

4. **Finite difference reduction formula; $\phi(x)$ a polynomial, unity.** For the formula about to be derived it is essential to distinguish between $\Delta\phi(x)/\Delta x$, which in this paper is written $[\phi(x)-\phi(x-\Delta x)]/\Delta x$, and $\{\Delta/\Delta x\}_{x_0}^x\phi(\xi)$ which, while the same as $\Delta\phi(x)/\Delta x$ for $x-x_0>\Delta x$, is but $\phi(x)/\Delta x$ for $x-x_0\leqq\Delta x$. While (4) would hold with $\{\Delta/\Delta x\}_{x_0}^x\phi(\xi)$ for $\{f_2(\Delta/\Delta x)\}_{x_0}^x\phi(\xi)$, it does not hold with $\Delta\phi(x)/\Delta x$ in place of $\{\Delta/\Delta x\}_{x_0}^x\phi(\xi)$. Instead, we obtain by the same method

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\frac{\Delta\phi(x)}{\Delta x}=\left\{\frac{\Delta}{\Delta x}f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\phi(x)-A\,[f(u)\,](p,\Delta x)\phi(x_0');$$

$$p=\left\{\frac{X-x_0}{\Delta x}\right\},\quad x_0'=X-(p+1)\Delta x.*$$

It will be noticed that $x_0'$ does not vary with $X$ as long as $X$ changes by multiples of $\Delta x$. Also $x_0-\Delta x<x_0'\leqq x_0$. If we replace $f(u)$ by $u^{-1}f(u)$ and rearrange its terms, this formula becomes

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\phi(x)=\left\{\left(\frac{\Delta}{\Delta x}\right)^{-1}f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\frac{\Delta\phi(x)}{\Delta x}+A\,[u^{-1}f(u)\,](p,\Delta x)\phi(x_0').$$

By induction, we are thus led to the fundamental formula

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\phi(x)=\left\{\left(\frac{\Delta}{\Delta x}\right)^{-m}f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X\frac{\Delta^m\phi(x)}{\Delta x^m}+\sum_{\mu=0}^{m-1}A\,[u^{-\mu-1}f(u)\,](p,\Delta x)\frac{\Delta^\mu\phi(x_0')}{\Delta x^\mu}.\tag{10}$$

It will be referred to as the finite difference reduction formula.

If $P(x)$ is a polynomial of degree $n$, we have

$$\frac{\Delta^{n+1}P(x)}{\Delta x^{n+1}}\equiv 0.$$

Let then $\phi(x)=P(x)$, $m=n+1$, in (10). It becomes

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^X P(x)=\sum_{\mu=0}^{n}A\,[u^{-\mu-1}f(u)\,](p,\Delta x)\frac{\Delta^\mu P(x_0')}{\Delta x^\mu}.$$

* The definition of $\{f(\Delta/\Delta x)\}_{x_0}^X\phi(x)$ only requires $\phi(x)$ to be defined in the interval $(x_0, X)$, whereas both members of this formula use values of $x$ less than $x_0$, when $(X-x_0)/\Delta x$ is not an integer. We must therefore arbitrarily define $\phi(x)$ for a suitable interval beyond $x_0$ to render the formula applicable. The validity of the formula, however, does not depend on the particular way in which this prolongation is effected.

If we now let $\Delta x \to +0$, it will follow that $p\Delta x \to X - x_0$, and $x_0' \to x_0$. Hence *if* $A[u^{-\mu-1}f(u)](X-x_0)$ *exists for* $\mu = 0, 1, \cdots, n$, $\{f(D)\}_{x_0}^{X}P(x)$ *will also exist, and will be given by*

$$\{f(D)\}_{x_0}^{X}P(x) = \sum_{\mu=0}^{n} A[u^{-\mu-1}f(u)](X - x_0)P^{(\mu)}(x_0). \tag{11}$$

For $P(x) \equiv 1$, we can use $n = 0$, and so obtain

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X} 1 = A[u^{-1}f(u)](p, \Delta x).$$

Hence, *if* $A[u^{-1}f(u)](X-x_0)$ *exists,* $\{f(D)\}_{x_0}^{X}1$ *exists, and is given by*

$$\{f(D)\}_{x_0}^{X}1 = A[u^{-1}f(u)](X - x_0). \tag{12}$$

It is easily verified that if $f(D)$ is of type zero then $D^{-\mu-1}f(D)$ also is of type zero. Hence the existence theorem for $\phi(x)$ a polynomial is completely established for operators of type zero.

The relation (12) is of special interest and importance. Note that in spite of the finite difference relation preceding it, we cannot conclude that if $\{f(D)\}_{x_0}^{X}1$ exists, $A[u^{-1}f(u)](X-x_0)$ also will exist; for in the former $p$ depends on $\Delta x$, while for the latter $p\Delta x$ should vary independently of $\Delta x$.

5. **Operators of finite order; existence theorems.** In establishing existence theorems for $\{f(D)\}_{x_0}^{X}\phi(x)$, we shall find that the wider the class of operators $f(D)$ we consider, the greater the restrictions we have to impose on $\phi(x)$. Hence the following specialization of operators of type zero.

An operator $f(D)$ will be said to be of *finite order* $\rho$, $\rho$ zero or a positive integer, when

(a) $f(z)$ *satisfies the analyticity condition for operators of type zero,*

(b) *there exists a positive* $\epsilon$, *and corresponding* $K$, *such that*

$$|f(z)| \leqq K|z|^{\rho-\epsilon}$$

*for every* $z$ *in the analytic sector for which* $|z| > \delta$, $\delta > 0$.

According to this definition, if an operator is of order $\rho$, it is also of any order greater than $\rho$.

$D^n$, $n$ arbitrary, is a typical example. All algebraic operators are of finite order. It is evident that operators of finite order are also of type zero. Hence §3 is immediately applicable.

We shall first derive an inequality for $A[f](r, \Delta x)$ which is essential for our existence proofs. With the notation of §3 we obtain by the new condition (b)

$$\left| \int_{C''} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < 2\pi K \left(\frac{k}{\Delta x}\right)^{\rho+1-\epsilon} / (k-1)^{r+1}, \quad k/\Delta x > \delta.$$

For fixed $\Delta x$, and $r \geqq \rho$, this expression approaches zero as limit as $k$ increases indefinitely, i.e., as the radius of $C''$ is made to increase indefinitely. We can therefore replace $C'''$ by $C$, the infinite contour of §3, and write, for $r \geqq \rho$,

$$A[f](r, \Delta x) = \frac{1}{2\pi i} \int_C \frac{f(z)dz}{(1 - z\Delta x)^{r+1}}.$$

Choose $C$ so that its vertex is at $z = 1/[2(r+1)\Delta x]$, and use condition (b) along it. This will be possible for $(r+1)\Delta x$ not too large. Through the change of variable $\zeta = z(r+1)\Delta x$, we then obtain

$$|A[f](r, \Delta x)| \leqq \frac{K}{2\pi} [(r+1)\Delta x]^{-\rho-1+\epsilon} \int_{C_{1/2}} \frac{|\zeta|^{\rho-\epsilon} |d\zeta|}{\left|1 - \frac{\zeta}{r+1}\right|^{r+1}},$$

where $C_{1/2}$ has its vertex at $\zeta = 1/2$. Since $R(\zeta) \leqq 1/2$ along $C_{1/2}$, we have

$$\frac{1}{\left|1 - \frac{\zeta}{r+1}\right|^{r+1}} \leqq \frac{1}{\left(1 - \frac{R(\zeta)}{r+1}\right)^{r+1}}.$$

Now we know that $[1 + x/(r+1)]^{r+1}$ is an increasing function of $r$, $r \geqq 0$, both for positive $x$, and for negative $x$ with $|x| < 1$. We therefore have along $C_{1/2}$

$$\frac{1}{\left(1 - \frac{R(\zeta)}{r+1}\right)^{r+1}} \leqq \frac{1}{\left(1 - \frac{R(\zeta)}{\rho+1}\right)^{\rho+1}}.$$

The integral along $C_{1/2}$, which depends only on $r$, is thus seen to be bounded for $r \geqq \rho$, so that we obtain

$$|A[f](r, \Delta x)| < L'[(r+1)\Delta x]^{-\rho-1+\epsilon}.$$

A like inequality is obtained for $r < \rho$ by using for $C'''$ a circle center $1/\Delta x$, radius $\theta/\Delta x$, $0 < \theta < 1$. We thus arrive at the following result:

*For all $\Delta x$'s and $r$'s with $(r+1)\Delta x < a$, where $a$ is some fixed positive quantity, we have the inequality*

$$|A[f](r, \Delta x)| < L[(r+1)\Delta x]^{-\rho-1+\epsilon}. \tag{13}$$

*Hence also, for $t < a$, we have*

$$|A[f](t)| \leqq Lt^{-\rho-1+\epsilon}. \tag{14}$$

The simplest formula for $\{f(D)\}_{x_0}^{X}\phi(x)$ results when $f(D)$ is of order zero. As at the end of §3, we shall write

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x) = \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x) + \sum_{r=q+1}^{p} A[f](r,\Delta x)\phi(X - r\Delta x)\Delta x;$$

$$q = \{h/\Delta x\}.$$

In the same manner also, we find for fixed positive $h$,

$$\lim_{\Delta x\to+0}\sum_{r=q+1}^{p} A[f](r,\Delta x)\phi(X - r\Delta x)\Delta x = \int_{x_0}^{X-h} A[f](X - x)\phi(x)dx.$$

Now $\phi(x)$ is to be assumed continuous in $(x_0, X)$. Let $M$ be the upper bound of $|\phi(x)|$ in this interval. For $h<a$ we can apply (13), with $\rho=0$, and obtain

$$\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x)\right| < LM\sum_{r=0}^{\{h/\Delta x\}}[(r+1)\Delta x]^{-1+\epsilon}\Delta x.$$

By comparing the indicated sum with the integral of $\xi^{-1+\epsilon}d\xi$ between limits 0 and $h+\Delta x$ or $\Delta x$ and $h+2\Delta x$, according as $0<\epsilon<1$, or $\epsilon\geqq 1$, we see that

$$\limsup_{\Delta x\to+0}\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x)\right| \leqq LM\frac{h^{\epsilon}}{\epsilon}.$$

Since $h^{\epsilon}/\epsilon\to 0$, as $h\to+0$, we can apply the limit criterion, and obtain

THEOREM III. *If $f(D)$ is of order zero, and $\phi(x)$ is continuous in $(x_0, X)$, then $\{f(D)\}_{x_0}^{X}\phi(x)$ exists, and is given by*

(15)

$$\{f(D)\}_{x_0}^{X}\phi(x) = \lim_{h\to+0}\int_{x_0}^{X-h} A[f](X - x)\phi(x)dx = \int_{x_0}^{X} A[f](X - x)\phi(x)dx.*$$

* If (8) is introduced in (15), we obtain

$$\{f(D)\}_{x_0}^{X}\phi(x) = \frac{1}{2\pi i}\int_{x_0}^{X}\left[\int_C e^{(X-x)z}f(z)dz\right]\phi(x)dx.$$

It may interest the reader to note that this formula might have suggested itself in the following formal manner. Symbolic use of Cauchy's second integral formula gives

$$f(D) = \frac{1}{2\pi i}\int_K \frac{f(z)dz}{z - D}.$$

On the other hand, the linear differential equation would suggest

$$\left\{\frac{1}{Z - D}\right\}_{x_0}^{X}\phi(x) = -\int_{x_0}^{X} e^{(X-x)z}\phi(x)dx,$$

so that we would be led to

$$\{f(D)\}_{x_0}^{X}\phi(x) = -\frac{1}{2\pi i}\int_K\left[\int_{x_0}^{X} e^{(X-x)z}\phi(x)dx\right]f(z)dz.$$

Reversing the sense in which "contour" $K$ is traversed, and changing the order of integration leads to the actual formula.

It will be noticed that the last integral may be improper for $x=X$; but of its convergence we are assured both by the limit criterion, and by the direct application of (14). To illustrate (15), we may take $f(D)=D^n$, $R(n)<0$. The operator is then of order zero. By (6) and (15), we thus get the standard Riemann-Grünwald form

$$\{D^n\}_{x_0}^{X}\phi(x) = \frac{1}{\Gamma(-n)}\int_{x_0}^{X}(X-x)^{-n-1}\phi(x)dx, \quad R(n)<0. \tag{16}$$

We shall give two existence proofs for operators of arbitrary finite order. The first proof depends on the observation that if $f(D)$ is of order $\rho$, then $D^{-\rho}f(D)$ is of order zero, and so comes under Theorem III. Inasmuch as additional assumptions on $\phi(x)$ in a left neighborhood of $x=X$ will be required, we shall use Theorem II with $x_1$ in this neighborhood.

Formula (10), with $x_1$ in place of $x_0$, and $m=\rho$, becomes

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\phi(x) = \left\{\left(\frac{\Delta}{\Delta x}\right)^{-\rho}f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\frac{\Delta^\rho\phi(x)}{\Delta x^\rho} + \sum_{\mu=0}^{\rho-1}A\,[u^{-\mu-1}f(u)\,](q,\Delta x)\frac{\Delta^\mu\phi(x_1')}{\Delta x^\mu}, \quad q=\{(X-x_1)/\Delta x\}.$$

Let $\phi(x)$ possess a continuous $\rho$th derivative in an interval $(k, X)$, $X>k$,* and assign to $x_1$ a value within this interval. By the law of the mean, and the continuity of $\phi^{(\rho)}(x)$ in $(k, X)$, it follows that, for sufficiently small $\Delta x$,

$$\left|\frac{\Delta^\rho\phi(x)}{\Delta x^\rho}-\phi^{(\rho)}(x)\right|<\delta_{\Delta x}, \quad \lim_{\Delta x\to+0}\delta_{\Delta x}=0,$$

for all $x$'s in $(x_1, X)$.† Hence if $\phi^{(\rho)}(x)$ be substituted for $\Delta^\rho\phi(x)/\Delta x^\rho$ in the above equation, the result will be changed in absolute value by no more than

$$\delta_{\Delta x}\sum_{r=0}^{q}|\,A\,[u^{-\rho}f(u)\,](r,\Delta x)\,|\,\Delta x.$$

Since $D^{-\rho}f(D)$ is of order zero, the proof of Theorem III shows this sum to be bounded, and hence the change in question to approach zero as limit as $\Delta x\to+0$. We then easily derive

* Here, as later, when a derivative is assumed to exist in a left neighborhood of $X$, the derivative at $X$ is to be but a left derivative.

† See de la Vallée Poussin, *Cours d'Analyse*, vol. 1, 1914, §109, for the case $\rho=1$. By (1), §118, this is directly extended to arbitrary $\rho$.

THEOREM IV. *If $f(D)$ is of order $\rho$, and if $\phi(x)$ is continuous in $(x_0, X)$, and possesses a continuous $\rho$th derivative in a left neighborhood of $x=X$, then $\{f(D)\}_{x_0}^{X}\phi(x)$ exists; and, if $x_1$ is within this neighborhood, is given by*

$$\{f(D)\}_{x_0}^{X}\phi(x) = \int_{x_1}^{X} A[u^{-\rho}f(u)](X-x)\phi^{(\rho)}(x)dx + \sum_{\mu=0}^{\rho-1} A[u^{-\mu-1}f(u)](X-x_1)\phi^{(\mu)}(x_1) + \int_{x_0}^{x_1} A[f](X-x)\phi(x)dx. \tag{17}$$

For illustration, again take $D^n$, with $\rho-1 \leqq R(n) < \rho$. We obtain

$$\{D^n\}_{x_0}^{X}\phi(x) = \frac{1}{\Gamma(\rho-n)}\int_{x_1}^{X}(X-x)^{\rho-n-1}\phi^{(\rho)}(x)dx + \sum_{\mu=0}^{\rho-1}\frac{(X-x_1)^{\mu-n}}{\Gamma(\mu-n+1)}\phi^{(\mu)}(x_1) + \frac{1}{\Gamma(-n)}\int_{x_0}^{x_1}(X-x)^{-n-1}\phi(x)dx.^{*} \tag{18}$$

For the second proof note that, except for successive operations, the upper limit $X$ plays the part of a constant. It can therefore appear as such in the operand $\phi(x)$. Now suppose that a continuous $\phi(x)$ satisfies the inequality

$$|\phi(x)| \leqq N(X-x)^{\rho}$$

in a left neighborhood of $x=X$. This, with (13), leads to the identical inequalities that gave us Theorem III. Hence (15) holds for such a $\phi(x)$.

Let then a continuous $\phi(x)$ possess a finite $\rho$th left derivative at $x=X$, and hence also left derivatives of all lower orders. Then, as a result of a first theorem in Taylor's expansion, we have with finite $N$, in a left neighborhood of $x=X$,

$$|\phi(x)-P(x)| < N(X-x)^{\rho}, \quad P(x) = \sum_{\mu=0}^{\rho-1}\frac{\phi^{(\mu)}(X)}{\mu!}(x-X)^{\mu}.$$

If $\phi(x)$ is continuous in $(x_0, X)$, $\{f(D)\}_{x_0}^{X}[\phi(x)-P(x)]$ will exist, as seen above, and be given by (15). On the other hand, $P(x)$ is but a polynomial in $x$, so that $\{f(D)\}_{x_0}^{X}P(x)$ exists, and could be evaluated by (11). Hence

THEOREM V. *If $f(D)$ is of order $\rho$, and if $\phi(x)$ is continuous in $(x_0, X)$, and possesses a finite $\rho$th left derivative at $x=X$, then $\{f(D)\}_{x_0}^{X}\phi(x)$ exists, and is given by*

$$\{f(D)\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X} A[f](X-x)\left[\phi(x) - \sum_{\mu=0}^{\rho-1}\frac{\phi^{(\mu)}(X)}{\mu!}(x-X)^{\mu}\right]dx + \{f(D)\}_{x_0}^{X}\sum_{\mu=0}^{\rho-1}\frac{\phi^{(\mu)}(X)}{\mu!}(x-X)^{\mu}. \tag{19}$$

* This reduces to Grünwald's result if $x_1=x_0$.

Though this theorem requires less of $\phi(x)$ than Theorem IV, we shall nevertheless find Theorem IV more useful in our development. To illustrate (19), we shall take the operator $\log D$, whose order is one. By direct calculation we have $A[\log u](t) = -1/t$. Hence, by (19) and (3), we get

$$(20)\quad \{\log D\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X} \frac{\phi(X) - \phi(x)}{X - x}dx - [\gamma + \log(X - x_0)]\phi(X).$$

It will be seen that this second proof uses the hypothesis that $f(D)$ is of finite order $\rho$ only to enable us to assume the existence of $A[u^{-\mu}f(u)](t)$, for $t>0$, $\mu=0, 1, \cdots, \rho$, and the validity of inequality (13). This suggests that we define $f(D)$ to be of *extended order* $\rho$ if $A[u^{-\mu}f(u)](t)$ exists for $t>0$, $\mu=0, 1, \cdots, \rho$, and (13) is satisfied. The first proof also uses the fact that $D^{-\rho}f(D)$ is of order zero. But it can be proved that if $f(D)$ is of extended order $\rho$, $D^{-\rho}f(D)$ is of extended order zero. Hence all of our existence theorems are valid for operators of extended finite order.

6. **Existence theorems for operators of type zero.** The crucial theorem for operators of type zero is the following:

THEOREM VI. *If $f(D)$ is of type zero, while $\phi(x)$ is analytic for $x_1 \leqq x \leqq X$, with the radius of convergence at $x_1$ greater than $X-x_1$, then $\{f(D)\}_{x_1}^{X}\phi(x)$ exists, and is given by the convergent series*

$$(21)\quad \{f(D)\}_{x_1}^{X}\phi(x) = A[u^{-1}f(u)](X-x_1)\phi(x_1) + A[u^{-2}f(u)](X-x_1)\phi'(x_1) + \cdots.$$

As a consequence of this theorem, if $\phi(x)$ is continuous in $(x_0, X)$, but analytic in some left neighborhood of $X$, then by choosing $x_1$ in this neighborhood, with $X-x_1$ less than half of the radius of convergence of $\phi(x)$ at $X$, both (9) and (21) become applicable, and so give

THEOREM VII. *If $f(D)$ is of type zero, while $\phi(x)$ is continuous in $(x_0, X)$, and analytic in a left neighborhood of $x=X$, then $\{f(D)\}_{x_0}^{X}\phi(x)$ exists, and, if $x_1$ is chosen as indicated above, is given by*

$$(22)\quad \{f(D)\}_{x_0}^{X}\phi(x) = \sum_{\mu=0}^{\infty} A[u^{-\mu-1}f(u)](X-x_1)\phi^{(\mu)}(x_1) + \int_{x_0}^{x_1} A[f](X-x)\phi(x)dx.$$

It may be noted that with 1 for $f(D)$, formula (21) reduces to the Taylor expansion of $\phi(x)$ for $x=x_1$. More generally, (21) is the result of operating on this Taylor expansion with $\{f(D)\}_{x_1}^{X}$, term by term.

Our proof of Theorem VI is an extension of the first existence proof for operators of finite order. As in that proof, we use the finite difference reduction formula, with limits $x_1$, $X$, viz.

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\phi(x) = \left\{\left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\frac{\Delta^m\phi(x)}{\Delta x^m} + \sum_{\mu=0}^{m-1} A\,[u^{-\mu-1}f(u)\,](q, \Delta x)\frac{\Delta^\mu\phi(x_1')}{\Delta x^\mu}.$$

On the other hand, we shall let $m$ vary with $\Delta x$, in fact equal $\{\epsilon/\Delta x\}$, with fixed and positive $\epsilon$. Our proof will require the following condition on $\epsilon$:

$$0 < \epsilon < \frac{l - (X - x_1)}{2},$$

where $l$ is between $X - x_1$ and the radius of convergence of $\phi(x)$ at $x_1$. Note that the left hand member of the reduction formula uses values of $x$ only in the interval $(x_1, X)$, whereas in the right hand member the values spread over the interval $(x_1 - m\Delta x, X)$. Since $m\Delta x < \epsilon < l$, all of these $x$'s fall in the interval of convergence of the Taylor expansion of $\phi(x)$ at $x_1$; and so this Taylor expansion may be used to define $\phi(x)$ for $x < x_1$ for the purpose of our proof.

This proof consists essentially in establishing the following:

(a) For fixed $N$, $\lim_{\Delta x \to +0} \sum_{\mu=0}^{N-1} A\,[u^{-\mu-1}f(u)\,](q, \Delta x)\frac{\Delta^\mu\phi(x_1')}{\Delta x^\mu} = \sum_{\mu=0}^{N-1} A\,[u^{-\mu-1}f(u)\,](X - x_1)\phi^{(\mu)}(x_1).$

(b) With $m = \{\epsilon/\Delta x\}$, $\lim_{\Delta x \to +0} \left\{\left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X} \frac{\Delta^m\phi(x)}{\Delta x^m} = 0.$

(c) With $m = \{\epsilon/\Delta x\}$, $\limsup_{\Delta x \to +0} \sum_{\mu=N}^{m-1} A\,[u^{-\mu-1}f(u)\,](q, \Delta x)\frac{\Delta^\mu\phi(x_1')}{\Delta x^\mu} \leqq h(N); \quad \lim_{N \to \infty} h(N) = 0.$

Since (b) allows us to neglect its term of the reduction formula, (a) and (c) together give us Theorem VI by means of the limit criterion.

(a) This is immediate.

(b) Formula (4) allows us to write

$$\left\{\left(\frac{\Delta}{\Delta x}\right)^{-m} f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\frac{\Delta^m\phi(x)}{\Delta x^m} = \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\left\{\left(\frac{\Delta}{\Delta \xi}\right)^{-m}\right\}_{x_1}^{x}\frac{\Delta^m\phi(\xi)}{\Delta \xi^m}, \quad \Delta\xi = \Delta x.$$

By the law of the mean we have

$$\frac{\Delta^m\phi(\xi)}{\Delta\xi^m} = \phi^{(m)}(\xi - \theta m\Delta x);\ 0 < \theta < 1.*$$

The Taylor expansion of $\phi(x)$ at $x_1$ enables us to define the function of a complex variable $\phi(z)$. Let $M$ be the upper bound of the modulus of $\phi(z)$ on, and hence also within, the circle center $x_1$, radius $l$. The smallest distance from $\xi-\theta m\Delta x$ to this circle is $l-|\xi-\theta m\Delta x-x_1|$, which certainly exceeds $l-\epsilon-(\xi-x_1)$ for all $\xi$'s in $(x_1, X)$. By applying the standard inequality of complex variable theory to $|\phi^{(m)}(\xi-\theta m\Delta x)|$, we thus obtain

$$\left|\frac{\Delta^m\phi(\xi)}{\Delta\xi^m}\right| < \frac{Mm!}{[l-\epsilon-(\xi-x_1)]^m},$$

so that we can write

$$\left|\left\{\left(\frac{\Delta}{\Delta\xi}\right)^{-m}\right\}_{x_1}^{x}\frac{\Delta^m\phi(\xi)}{\Delta\xi^m}\right| < \sum_{s=0}^{[(x-x_1)/\Delta x]} A[u^{-m}](s,\Delta x)\frac{Mm!}{[l-\epsilon-(x-s\Delta x-x_1)]^m}\Delta x.$$

It is easily seen that

$$A[u^{-m}](s,\Delta x) = \frac{m(m+1)\cdots(m+s-1)}{s!}\Delta x^{m-1} < \frac{(\epsilon+s\Delta x)^{m-1}}{(m-1)!}.$$

Now the inequalities imposed on $\epsilon$ make $(\epsilon+s\Delta x)/[l-\epsilon-(x-s\Delta x-x_1)]$ an increasing function of $s\Delta x$. Since $s\Delta x<x-x_1$, we thereby obtain, on replacing $s\Delta x$ by $x-x_1$, the inequality

$$\frac{\epsilon+s\Delta x}{l-\epsilon-(x-s\Delta x-x_1)} < \frac{\epsilon+(x-x_1)}{l-\epsilon},$$

and so find in succession

$$\left|\left\{\left(\frac{\Delta}{\Delta\xi}\right)^{-m}\right\}_{x_1}^{x}\frac{\Delta^m\phi(\xi)}{\Delta\xi^m}\right| < \frac{M(x-x_1)}{\epsilon}\cdot m\left[\frac{\epsilon+(x-x_1)}{l-\epsilon}\right]^m,$$

$$\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_1}^{X}\left\{\left(\frac{\Delta}{\Delta\xi}\right)^{-m}\right\}_{x_1}^{x}\frac{\Delta^m\phi(\xi)}{\Delta\xi^m}\right| < \frac{M(X-x_1)}{\epsilon}\cdot m\left[\frac{\epsilon+(X-x_1)}{l-\epsilon}\right]^m\sum_{r=0}^{q}|A[f](r,\Delta x)|\,\Delta x.$$

* This, as also the use of the law of the mean in §5, requires $\phi(x)$ to be a real function of the real variable $x$. If $\phi(x)$ is a complex function of a real variable, its real and imaginary parts separately will satisfy the assumption of existence and continuity of derivatives or of analyticity stated for $\phi(x)$ so that the demonstrations given are valid for these parts. Combining the two results thus obtained therefore yields the same result for the complex $\phi(x)$.

If $\delta$ is a real fixed quantity in the analytic sector of $f(z)$, we can choose for the $C'''$ of §3 a circle, center $1/\Delta x$, radius $1/\Delta x - \delta$, and so, through the corresponding contour integral and condition (b) of §3, obtain

$$|\Delta[f](r, \Delta x)| < \frac{Ke^{\kappa(2/\Delta x-\delta)}}{(1-\delta\Delta x)^r}.$$

$\sum_{\mu=0}^{q} |\Delta[f](r, \Delta x)|\Delta x$ is then less than $e^{2\kappa/\Delta x}$ times the sum of a geometric progression which is easily seen to have a finite limit as $\Delta x \to +0$. The essential factor is thus $e^{2\kappa/\Delta x}$. Now the inequalities imposed on $\epsilon$ can be rewritten

$$0 < \frac{\epsilon + (X - x_1)}{l - \epsilon} < 1.$$

Hence, by choosing $\kappa$ sufficiently small, we obtain

$$\lim_{\Delta x \to +0} m\left[\frac{\epsilon + (X - x_1)}{l - \epsilon}\right]^m e^{2\kappa/\Delta x} = 0, \quad m = \left\{\frac{\epsilon}{\Delta x}\right\},$$

thereby establishing (b).

(c) As in part (b), we obtain

$$\left|\frac{\Delta^\mu \phi(x_1')}{\Delta x^\mu}\right| < \frac{M\mu!}{(l-\epsilon)^\mu}.$$

On the other hand

$$\Delta[u^{-\mu-1}f(u)](q, \Delta x) = -\frac{1}{2\pi i}\int_W \frac{z^{-\mu-1}f(z)}{(1 - z\Delta x)^{q+1}}dz,$$

where $W$ encloses $1/\Delta x$, but excludes the origin. The contour $W$ will consist of $W_1$, that part of the $C'''$ of §3 for which $|z| > R$, $R < 1/\Delta x$, joined to $W_2$, the arc of the circle center the origin, radius $R$, cut off by $C'''$, and excluding, with $W_1$, the origin. Such a contour will be valid when $R$ exceeds the distance of the vertex of $C$ from the origin. We have

$$\left|\frac{1}{2\pi}\int_{W_2} \frac{z^{-\mu-1}f(z)}{(1 - z\Delta x)^{q+1}}dz\right| < K\frac{R^{-\mu}e^{\kappa R}}{(1 - R\Delta x)^{q+1}}.$$

Now $R^{-\mu}/(1-R\Delta x)^{q+1}$ has a minimum with respect to $R$ for

$$R = \frac{\mu}{(q+1)\Delta x + \mu\Delta x}.$$

This value of $R$ is less than $1/\Delta x$, and exceeds $\mu/[(X-x_1)+\epsilon]$, so that, for sufficiently large $\mu$, it will give a valid contour. With it we have

$$\frac{R^{-\mu}e^{\kappa R}}{(1-R\Delta x)^{q+1}} = \frac{[(q+1)\Delta x+\mu\Delta x]^{\mu}}{\mu^{\mu}} e^{\kappa\mu/[(q+1)\Delta x+\mu\Delta x]}\left(1+\frac{\mu}{q+1}\right)^{q+1},$$

and, as $\mu<m$, we obtain

$$\left|\frac{1}{2\pi}\int_{W_2}\frac{z^{-\mu-1}f(z)}{(1-z\Delta x)^{q+1}}dz\right| < K[(X-x_1)+\epsilon]^{\mu}e^{\kappa\mu/(X-x_1)}\mu^{-\mu}e^{\mu}.$$

For $W_1$ we have $|z|>R$. Since $W_1$ is part of $C'''$, we easily find that

$$\left|\frac{1}{2\pi}\int_{W_1}\frac{z^{-\mu-1}f(z)}{(1-z\Delta x)^{q+1}}dz\right| < \frac{[(X-x_1)+\epsilon]^{\mu+1}}{\mu^{\mu+1}}\cdot\frac{1}{2\pi}\int_{C'''}\frac{|f(z)|}{|1-z\Delta x|^{q+1}}|dz|.$$

Now our discussion of $A[f](r,\Delta x)$ in §3 proves that this integral has a finite limit as $\Delta x\to+0$. Symbolizing the product of $1/(2\pi)$ and this limit by $||A[f](X-x_1)||$, we are thus led to the following form for the $h(N)$ of the statement of this part of the proof, viz.,

$$h(N) = MK\sum_{\mu=N}^{\infty}\left\{\left[\frac{(X-x_1)+\epsilon}{l-\epsilon}\right]^{\mu}e^{\kappa\mu/(X-x_1)}\frac{\mu!}{\mu^{\mu}e^{-\mu}}\right\}$$
$$+ M(l-\epsilon)||A[f](X-x_1)||\sum_{\mu=N}^{\infty}\left\{\left[\frac{(X-x_1)+\epsilon}{l-\epsilon}\right]^{\mu+1}\frac{\mu!}{\mu^{\mu+1}}\right\}.$$

Recalling that we have $0<[(X-x_1)+\epsilon]/(l-\epsilon)<1$, we see that the second series converges. By choosing $\kappa$ sufficiently small to have

$$\frac{(X-x_1)+\epsilon}{l-\epsilon}e^{\kappa/(X-x_1)} < 1,$$

the convergence of the first series is assured. As a consequence of their convergence, these series approach zero as limit as $N$ increases indefinitely. The same is therefore true of $h(N)$. (c) has thus been proved, and with it Theorem VI.

We turn now to the formal properties mentioned in the introduction.

7. **Differentiation with respect to the upper limit.** We can obtain the derivative of $A[f](t)$ under a general hypothesis. If $A[f(u)](t)$ and $A[uf(u)](t)$ exist for $0<t_1\leqq t<t_2$,* then the terms of the relation

$$\{Df(D)\}_{-t}^{0}1 = \{Df(D)\}_{-t_1}^{0}1 + \int_{-t}^{-t_1}A[uf(u)](-\xi)d\xi$$

* The relation given below shows that if $A[uf(u)](t)$ exists in a neighborhood, and $A[f(u)](t)$ exists for one $t$ in that neighborhood, it exists for every $t$ therein.

obtained from Theorem II, exist, since by (12) this relation becomes

$$A[f(u)](t) = A[f(u)](t_1) + \int_{t_1}^{t} A[uf(u)](\eta)d\eta.$$

By keeping $t_1$ constant, and differentiating with respect to $t$, we get

THEOREM VIII. *If $A[f(u)](t)$ and $A[uf(u)](t)$ exist in a certain neighborhood of $t$, then $(d/dt)\,A[f(u)](t)$ exists, and is given by*

$$\frac{d}{dt}A[f(u)](t) = A[uf(u)](t). \tag{23}$$

By the use of (9), we immediately obtain the

COROLLARY. *If $A[u^{-1}f(u)](X-x_0)$ and $A[f(u)](X-x_0)$ exist in a certain neighborhood of $X$, then $(d/dX)\ \{f(D)\}_{x_0}^{X}1$ exists, and is given by*

$$\frac{d}{dX}\{f(D)\}_{x_0}^{X}1 = \{Df(D)\}_{x_0}^{X}1. \tag{24}$$

In extending this corollary to $\phi(x)$ as operand, it is desirable to replace the $X$ left neighborhood of our existence theorems by a complete neighborhood, so that $d/dX$ can stand for the derivative as ordinarily used. If, however, we retain the left neighborhood, then the following still remains valid, provided $d/dX$ is understood to mean left derivative, an observation that is essential to most of our applications of these formulas.

For $f(D)$ either of finite order, or of type zero, we have, as a result of(23),

$$\frac{d}{dX}\int_{x_0}^{x_1} A[f(u)](X-x)\phi(x)dx = \int_{x_0}^{x_1} A[uf(u)](X-x)\phi(x)dx,$$

since the continuity of $A[uf(u)](X-x)$, for $x_0 \leqq x \leqq x_1$, and a neighborhood of $X$, permits differentiation under the integral sign. Hence, by (9), we shall have $(d/dX)\{f(D)\}_{x_0}^{X}\phi(x) = \{Df(D)\}_{x_0}^{X}\phi(x)$, provided we first prove this relation for lower limit $x_1$. We shall choose $x_1$, so that the special expansions hold.

First let $Df(D)$ be of order zero. $f(D)$ is then also of order zero, so that we can use (15), and write

$$\{f(D)\}_{x_1}^{X}\phi(x) = \lim_{h\to+0}\int_{x_1}^{X-h} A[f(u)](X-x)\phi(x)dx.$$

By Leibnitz's rule we have

$$\frac{d}{dX}\int_{x_1}^{X-h} A[f(u)](X-x)\phi(x)dx = A[f(u)](h)\phi(X-h) + \int_{x_1}^{X-h} A[uf(u)](X-x)\phi(x)dx.$$

Furthermore, by (14), we have

$$|A[uf(u)](X-x)| \leqq L(X-x)^{-1+\epsilon}, \quad |A[f(u)](h)| \leqq L'h^{\epsilon *}.$$

These inequalities show that the result of the last differentiation uniformly approaches $\int_{x_1}^{X} A[uf(u)](X-x)\phi(x)dx$ as limit in a neighborhood of $X$, as $h \to +0$. Hence by a well known criterion of differentiation,† we have, as desired,

$$\frac{d}{dX}\{f(D)\}_{x_1}^{X}\phi(x) = \int_{x_1}^{X} A[uf(u)](X-x)\phi(x)dx.$$

If now $Df(D)$ is of order $\rho$, we can use the first existence theorem for operators of finite order with $f(D)$ as of order $\rho$, and write

$$\{f(D)\}_{x_1}^{X}\phi(x) = \{D^{-\rho}f(D)\}_{x_1}^{X}\phi^{(\rho)}(x) + \sum_{\mu=0}^{\rho-1} A[u^{-\mu-1}f(u)](X-x_1)\phi^{(\mu)}(x_1).$$

Since $D \cdot D^{-\rho}f(D)$ is of order zero, we can use the special case just proved, along with (23), in differentiating both members of this equation, thereby obtaining on the right the expansion of $\{Df(D)\}_{x_1}^{X}\phi(x)$. Hence

THEOREM IX. *If $Df(D)$ is of order $\rho$, and if $\phi(x)$ is continuous in $(x_0, X)$, and has a continuous $\rho$th derivative in a neighborhood of $X$, then*

$$\frac{d}{dX}\{f(D)\}_{x_0}^{X}\phi(x) = \{Df(D)\}_{x_0}^{X}\phi(x). \tag{25}$$

COROLLARY. *If $D^n f(D)$ is of order $\rho$, and if $\phi(x)$ is continuous in $(x_0, X)$, and has a continuous $\rho$th derivative in a neighborhood of $X$, then*

$$\frac{d^n}{dX^n}\{f(D)\}_{x_0}^{X}\phi(x) = \{D^n f(D)\}_{x_0}^{X}\phi(x). \tag{26}$$

If $f(D)$ is replaced by $D^{-\rho}f(D)$, and $n$ by $\rho$, the hypothesis of this corollary, restricted to a left neighborhood of $X$, reduces to the hypothesis of Theorem IV. Under this hypothesis, (26) therefore reduces to

---

* The first because $\rho=0$ for $uf(u)$; the second because the $\rho$ of $f(u)$ may be taken to be $-1$. Though $\rho$, as defined in §5, is either zero or a positive integer, it can be assigned a negative value, as is convenient here, with (14) remaining valid, provided a positive $\epsilon$ can still be chosen.

† Goursat, *Cours d'Analyse*, vol. 1, 1910, p. 74.

$$\{f(D)\}_{x_0}^{X}\phi(x) = \frac{d^\rho}{dX^\rho}\int_{x_0}^{X} A\,[u^{-\rho}f(u)\,](X-x)\phi(x)dx. \tag{27}$$

Formula (27) is a direct extension of Riemann's form for $D^n$ for $R(n)>0$.

The extension of Theorem IX to operators of type zero offers no difficulty. We first observe that the result of differentiating series (21) term by term is the expansion of $\{Df(D)\}_{x_1}^{X}\phi(x)$. Hence it is only necessary to prove the resulting series uniformly convergent in a neighborhood of $X$ to obtain the desired result. Since $Df(D)$ is also of type zero, no loss of generality ensues if we prove the series for $\{f(D)\}_{x_1}^{X}\phi(x)$ uniformly convergent. Turning to the discussion of (c) §6, we observe that the single series in which $h(N)$ can be written is a majorant for the corresponding part of the series for $\{f(D)\}_{x_1}^{X}\phi(x)$. Now the discussion of §3 shows that we can write

$$||\,A\,[f](X-x_1)\,|| = \frac{1}{2\pi}\int_C e^{R(z)(X-x_1)}\,|\,f(z)\,|\,\,|\,dz\,|\,.$$

This integral is continuous in $X$, and hence bounded for the neighborhood of $X$ in question. On the other hand, through the choice of $\kappa$ made in (c) §6, we shall have, for a sufficiently small neighborhood of $X$,

$$0 < \frac{(X-x_1)+\epsilon}{l-\epsilon}e^{\kappa/(X-x_1)} < \lambda < 1.$$

As a result, the terms of $\{f(D)\}_{x_1}^{X}\phi(x)$ will be less in absolute value than those of a convergent series of positive constants, and so the series is uniformly convergent. We thus have

THEOREM X. *If $f(D)$ is of type zero, and $\phi(x)$ satisfies the hypothesis of Theorem* VII, *extended to a complete neighborhood of $X$, then $(d/dX)\{f(D)\}_{x_0}^{X}\phi(x)$ exists, and is given by* (25).

8. **The law of successive operations.** In the present section we shall consider certain conditions under which the relation

$$\{f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi) = \{f_1(D)f_2(D)\}_{x_0}^{X}\phi(x) \tag{28}$$

is valid. Inasmuch as $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$, which serves as operand for $f_1(D)$, may be discontinuous at $x=x_0$, we shall introduce the following extension of our fundamental definition. If $\phi(x)$ is discontinuous for $x=x_0$, while $\{f(D)\}_{x_0+h}^{X}\phi(x)$ exists for sufficiently small $h$, then $\{f(D)\}_{x_0}^{X}\phi(x)$ is to be defined by

$$\{f(D)\}_{x_0}^{X}\phi(x) = \lim_{h\to+0}\{f(D)\}_{x_0+h}^{X}\phi(x) \tag{29}$$

provided the limit in question exists. Under the hypothesis of Theorem II, changed to allow for a discontinuity of $\phi(x)$ for $x=x_0$, the existence of this limit is equivalent to the convergence of the improper integral thus occurring in formula (9). This convergence is amply assured if $\phi(x)$ satisfies in a right neighborhood of $x_0$ the inequality

$$|\phi(x)| \leqq M(x - x_0)^{-1+\eta}, \ \eta > 0. \tag{30}$$

It will be convenient to say that $\phi(x)$ then has $M(x-x_0)^{-1+\eta}$ as majorant in the neighborhood. With (30) to replace continuity of $\phi(x)$ at $x=x_0$, Theorem II, with formula (9), continues to hold. Our existence theorems therefore go over, as do also the results of the preceding section.*

Consider, however, the special case $f_2(D)=D^n$, $n$ other than zero, or a positive integer. We have, by (12) and (6),

$$\{D^n\}_{x_0}^{x} 1 = A[u^{n-1}](x - x_0) = (x - x_0)^{-n}/\Gamma(-n+1).$$

This is discontinuous for $x=x_0$ when $R(n)>0$, so that, without the above extension of our definition, (28) could only hold for $R(n)<0$. With this extension, we can have $R(n)<1$, since $\{D^n\}_{x_0}^{x}1$ then satisfies (30). On the other hand, for $R(n)>1$, let $f_1(D)=D^{-1}$. We have

$$\{D^{-1}\}_{x_0+h}^{X}\{D^n\}_{x_0}^{x} 1 = \int_{x_0+h}^{X} \frac{(x - x_0)^{-n}}{\Gamma(-n+1)} dx,$$

which diverges as $h\to+0$. The left hand member of (28) therefore fails to exist in this case.

Volterra has encountered the same difficulty in the related theory of functions of composition; but his solution appears to the writer to be but a verbal evasion.† Among other possibilities, the difficulty might be removed by a study of the commonly neglected arbitrary series that Riemann adds to the definite integral in his formula for $D^n$. In the absence of a definitive solution, the writer leaves the breach open to view.‡ This possible failure

* The differentiation of an improper integral under the integral sign that is required here is easily justified by the criterion of differentiation referred to in §7.

† V. Volterra, *Functions of composition*, The Rice Institute Pamphlet, vol. 7 (1920), p. 202.

‡ This difficulty does not appear in the theory for infinite lower limit; but the validity of our definition for finite lower limit is thereby rendered questionable. It is to be noted, however, that this failure occurs in the first place for the commonly accepted Riemann-Grünwald form for $D^n$. Furthermore, in all other respects the theory for finite lower limit is satisfactory; and if not for its own value, it would still be required as a foundation for the theory for infinite lower limit. Finally, in the various tentative modifications of the definition considered by the writer, the present theory remains the indispensable basis for the extension.

of (28), because of the non-existence of the second of the two successive operations, forces us to restrict $f_2(D)$ to finite orders zero or one, the latter only made possible by the extension of definition just given. Because of this restriction of $f_2(D)$, we shall only consider $f_1(D)$'s of finite order.

Our first consideration is to find conditions on $\phi(x)$ which will insure the existence of the left hand member of (28). For this purpose $f_1(D)$ may be taken of order zero, the extension to arbitrary finite order being easily made. We shall then want $\{f_2(D)\}_{x_0}\phi(\xi)$ to exist and be continuous for $x_0<x\leqq X$, and to satisfy in a right neighborhood of $x_0$ an inequality of type (30).

First let $f_2(D)$ be of order zero. Then, to insure the existence of $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$, we shall want $\phi(\xi)$ to be continuous for $x_0<\xi\leqq X$, and to have $M(\xi-x_0)^{-1+\eta}$, $\eta>0$, as majorant in a right neighborhood of $x_0$. The same conditions turn out to be sufficient to yield the remaining requirements for $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$. In fact, note that, due to its continuity, $\phi(\xi)$ will have $M'(\xi-x_0)^{-1+\eta}$ as majorant over the whole interval $(x_0, X)$. By applying (15), and using $Lt^{-1+\epsilon}$ as majorant for $A[f_2](t)$, in accordance with (14), we easily establish the continuity of $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$. Furthermore, by the substitution of majorants, and reduction to the first Eulerian integral, (15) yields

$$|\{f_2(D)\}_{x_0}^{x}\phi(\xi)| \leqq LM'B(\epsilon, \eta)(x-x_0)^{-1+\epsilon+\eta},$$

which is of type (30).

When $f_2(D)$ is of order one, the use of the first existence theorem for operators of finite order requires $\phi(\xi)$ to possess a continuous first derivative for $x_0<\xi\leqq X$, since $x$ must vary from $x_0$ to $X$. This time, $\phi'(\xi)$ is assumed to have $M(\xi-x_0)^{-2+\eta}$, $\eta>0$, as majorant. Since no loss of generality ensues if $\eta$ is assumed less than one, $\phi(\xi)$ will then have $M'(\xi-x_0)^{-1+\eta}$ as majorant. By applying (17) with $x-x_1=(x-x_0)/2$, i.e. by writing

$$\{f_2(D)\}_{x_0}^{x}\phi(\xi) = \int_{x_1}^{x} A[u^{-1}f(u)](x-\xi)\phi'(\xi)d\xi + A[u^{-1}f_2(u)](x-x_1)\phi(x_1) + \int_{x_0}^{x_1} A[f_2(u)](x-\xi)\phi(\xi)d\xi,$$

and using the majorants $Lt^{-2+\epsilon}$ and $L't^{-1+\eta}$ for $A[f_2](t)$ and $A[u^{-1}f_2(u)](t)$ respectively, as given by (14), in conjunction with the above majorants for $\phi'(\xi)$ and $\phi(\xi)$, we establish the stated requirements for $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$, under the added condition $\epsilon+\eta>1$.

The desired conditions are therefore the following:

(a) for $f_2(D)$ of order zero: $\phi(x)$ is continuous for $x_0<x\leqq X$, with $M(x-x_0)^{-1+\eta}$, $\eta>0$, as majorant;

(b) for $f_2(D)$ of order one: $\phi'(x)$ exists and is continuous for $x_0<x\leqq X$, with $M(x-x_0)^{-2+\eta}$ as majorant, where $\eta>0$, and $\epsilon+\eta>1$.

We proceed now to establish (28) for $f_1(D)$ of order zero under the above conditions on $\phi(x)$. The following preliminary result is fundamental.

As in the derivation of (4) §1, we find

$$(31)\quad A[f_1(u)f_2(u)](p,\Delta x)=\sum_{r=0}^{p}A[f_1(u)](r,\Delta x)A[f_2(u)](p-r,\Delta x)\Delta x.$$

If we let $p=\{(X-x_0)/\Delta x\}$, and have $\Delta x\to+0$, then, by a slight modification of the method used in establishing (15), we obtain

$$(32)\quad A[f_1(u)f_2(u)](X-x_0)=\int_{x_0}^{X}A[f_1(u)](X-x)A[f_2(u)](x-x_0)dx,^*$$

provided $f_1(D)$ and $f_2(D)$ are both of order zero. By (15) this becomes

$$(33)\qquad \{f_1(D)\}_{x_0}^{X}A[f_2(u)](x-x_0)=A[f_1(u)f_2(u)](X-x_0).$$

Let then $f_1(D)$ and $f_2(D)$ first both be of order zero, with $\phi(x)$ satisfying condition (a) given above. We can in this case apply (15), and write

$$\{f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi)=\int_{x_0}^{X}A[f_1(u)](X-x)\int_{x_0}^{x}A[f_2(u)](x-\xi)\phi(\xi)d\xi dx.$$

The improper double integral corresponding to this iterated integral exists, since the latter, and hence the former, is absolutely convergent.† Consequently the order of integration can be changed to give

$$\{f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi)=\int_{x_0}^{X}\left[\int_{\xi}^{X}A[f_1(u)](X-x)A[f_2(u)](x-\xi)dx\right]\phi(\xi)d\xi.$$

By (32) and (15) this reduces to the desired relation (28).

Now let $f_1(D)$ be of order zero, $f_2(D)$ of order one, with $\phi(x)$ satisfying condition (b). Let $x_1$ be chosen between $x_0$ and $X$. We have, by Theorem IV,

$$\{f_1(D)\}_{x_1}^{X}\{f_2(D)\}_{x_1}^{x}\phi(\xi)=\{f_1(D)\}_{x_1}^{X}\{D^{-1}f_2(D)\}_{x_1}^{x}\phi'(\xi)+\phi(x_1)\{f_1(D)\}_{x_1}^{X}A[u^{-1}f_2(u)](x-x_1).$$

* This formula is of special interest in connection with Volterra's functions of composition of the closed cycle group. (See V. Volterra, loc. cit., pp. 181–251.) If $A[f_1(u)](X-x)$ is written $g_1(X-x)$, and $A[f_2(u)](x-x_0)$, $g_2(x-x_0)$, then (32) shows $A[f_1(u)f_2(u)](X-x_0)$ to be Volterra's $g_1^*g_2$, so that this symbolic product of the $g$'s corresponds to the actual product of the related $f$'s.

† See de la Vallée Poussin, *Cours d'Analyse*, vol. 2, 1925, pp. 19–22.

Since $D^{-1}f_2(D)$ is of order zero, (28) can be applied to the first term, and (33) to the second term, of the right hand member to give

$$\{f_1(D)\}_{x_1}^{X}\{f_2(D)\}_{x_1}^{x}\phi(\xi) = \{D^{-1}f_1(D)f_2(D)\}_{x_1}^{X}\phi'(x)+A[u^{-1}f_1(u)f_2(u)](X-x_1)\phi(x_1)$$
$$= \{f_1(D)f_2(D)\}_{x_1}^{X}\phi(x).$$

(29) then shows that to demonstrate (28) in this case, we need but establish the relation

$$\lim_{x_1\to x_0}\{f_1(D)\}_{x_1}^{X}\{f_2(D)\}_{x_1}^{x}\phi(\xi) = \{f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi).$$

Now we have directly

$$\{f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi) - \{f_1(D)\}_{x_1}^{X}\{f_2(D)\}_{x_1}^{X}\phi(\xi)$$
$$= \int_{x_0}^{x_1} A[f_1(u)](X-x)\left[\{f_2(D)\}_{x_0}^{x}\phi(\xi)\right]dx$$
$$+ \{f_1(D)\}_{x_1}^{X}\left[\int_{x_0}^{x_1} A[f_2(u)](x-\xi)\phi(\xi)d\xi\right].$$

The first term of the second member of this equation is immediately seen to approach zero as limit as $x_1\to x_0$. As for the second term, return to the discussion of the conditions imposed on $\phi(x)$. We can assume without loss of generality the inequalities $0<\epsilon<1$, $0<\eta<1$. Since $\epsilon+\eta>1$ we can choose a positive $\lambda$ such that $\lambda<\epsilon+\eta-1$. We therefore also have $\lambda<\epsilon$. Then in connection with the integral appearing in the second term, we have for the corresponding majorants

$$\int_{x_0}^{x_1}(x-\xi)^{-2+\epsilon}(\xi-x_0)^{-1+\eta}d\xi < (x-x_1)^{-1+\lambda}\int_{x_0}^{x_1}(x_1-\xi)^{-1+\epsilon-\lambda}(\xi-x_0)^{-1+\eta}d\xi$$
$$= (x-x_1)^{-1+\lambda}(x_1-x_0)^{\epsilon+\eta-1-\lambda}B(\epsilon-\lambda,\eta).$$

This second term will therefore be less in absolute value than

$$M''(X-x_1)^{\mu}(x_1-x_0)^{\epsilon+\eta-1-\lambda}.$$

Since $\epsilon+\eta-1-\lambda>0$, this term also approaches zero as limit as $x_1\to x_0$. The desired relation is thus established, and with it, (28).

The extension to $f_1(D)$ of arbitrary finite order $\rho_1$, with $f_2(D)$ of order zero or one, is now easily made. $D^{-\rho_1}f_1(D)$ will be of order zero, so that, under the above conditions on $\phi(x)$,

$$\{D^{-\rho_1}f_1(D)\}_{x_0}^{X}\{f_2(D)\}_{x_0}^{x}\phi(\xi) = \{D^{-\rho_1}f_1(D)f_2(D)\}_{x_0}^{X}\phi(x).$$

We have observed that the results of §7 remain valid under (29) and (30). Note also that the derivative appearing in Theorem IX and Corollary I is continuous in a left neighborhood of $x=X$. Suppose then that $D^{\rho_1}f_2(D)$ is of order $\rho$, and that $\phi(x)$ possesses a continuous $\rho$th derivative in a left neighborhood of $x=X$. Then $\{f_2(D)\}_{x_0}^{x}\phi(\xi)$ will possess a continuous $\rho_1$th derivative in a left neighborhood of $x=X$. Now $D^{\rho_1}\cdot D^{-\rho_1}f_1(D)$ is of order $\rho_1$; $D^{\rho_1}\cdot D^{-\rho_1}f_1(D)f_2(D)$ is of order not greater than $\rho$. Hence we can operate on both sides of the above equation by $d^{\rho_1}/dX^{\rho_1}$ in accordance with the corollary of Theorem IX, and by (26) obtain (28).

These results can be stated as

THEOREM XI. *If $f_1(D)$ and $f_2(D)$ are of orders $\rho_1$ and $\rho_2$ respectively, with $\rho_2$ equal to zero or one, while $D^{\rho_1}f_2(D)$ is of order $\rho$; and if $\phi(x)$ has a continuous $\rho$th derivative in a left neighborhood of $x=X$, a continuous $\rho_2$th derivative for $x_0<x\leqq X$, while for this interval $|\phi^{(\rho_2)}(x)|\leqq M(x-x_0)^{-\rho_2-1+\eta}$, where $\eta$ is positive and such that $D^{-\eta}f_2(D)$ is of order zero,* then* (28) *is valid.*

The restriction of $f_2(D)$ to order zero or one was necessary in order that the result should apply to $\phi(x)\equiv 1$, the simplest case. However, by requiring $\phi(x)$ and a sufficient number of its derivatives to vanish for $x=x_0$, the order of $f_2(D)$ can be arbitrarily large. In fact under the specific conditions stated below, formula (17) gives

$$\{f_2(D)\}_{x_0}^{x}\phi(\xi) = \{D^{-\rho_2+1}f_2(D)\}_{x_0}^{x}\phi^{(\rho_2-1)}(\xi).$$

As $D^{-\rho_2+1}f_2(D)$ will then be of order one, the theorem applies. Hence the

COROLLARY. *The relation* (28) *holds for $\rho_2>1$, provided the corresponding conditions of the theorem are replaced by the following: $\phi^{(\rho_2-2)}(x)$ exists, and is continuous, for $x_0\leqq x\leqq X$, with $\phi^{(\mu)}(x_0)=0$, for $\mu=0, 1, \cdots, \rho_2-2$; $\phi^{(\rho_2)}(x)$ exists, and is continuous, for $x_0<x\leqq X$, with $|\phi^{(\rho_2)}(x)|\leqq M(x-x_0)^{-2+\eta}$, where $\eta$ is positive, and such that $D^{-\eta}f_2(D)$ is of order $\rho_2-1$.*

The reader can apply these results to more than two successive operations, and to the commutativity of the operators.

When either of the $f$'s is a polynomial, no restriction of the order of $f_2(D)$ is needed. In fact, if $P(d/dx)$ is a polynomial in $d/dx$ we obtain, by (26),

$$P\left(\frac{d}{dX}\right)\{f(D)\}_{x_0}^{X}\phi(x) = \{P(D)f(D)\}_{x_0}^{X}\phi(x), \tag{34}$$

under the condition that, if $P(D)f(D)$ is of order $\rho$, $\phi(x)$ satisfies the hypothesis of Theorem IX. On the other hand, through (17), we get

* This is another way of stating the condition $\epsilon+\eta>1$ used in the proof.

$$(35)\qquad \{f(D)\}_{x_0}^{X} P\left(\frac{d}{dx}\right)\phi(x) = \{P(D)f(D)\}_{x_0}^{X}\phi(x) - \sum_{\mu=0}^{n-1} A[u^{n-\mu-1}f(u)](X-x_0)P_\mu(x_0),$$

where

$$P(D) = \sum_{\nu=0}^{n} a_\nu D^{n-\nu},\quad P_\mu(x_0) = \sum_{\nu=0}^{\mu} a_\nu \phi^{(\mu-\nu)}(x_0).$$

For (35), $\phi(x)$ is to possess a continuous $n$th derivative in $(x_0, X)$, and a continuous $\rho$th derivative in a left neighborhood of $X$, where $\rho$ is the order of $P(D)f(D)$. Formulas (34) and (35) can easily be extended to the case where $f(D)$ is of type zero. They probably also admit of extension to the case where $P(D)$ is not a polynomial, but of type zero over the plane. (See §14.) Though the disagreement between (35) and (28) was to be expected as a result of the finite difference reduction formula, it may suggest a solution of the difficulty discussed above.

9. **Leibnitz's theorem generalized.** The operators $f^n(D)$, $n=1, 2, 3, \cdots$, are intimately associated with the generalization of Leibnitz's Theorem for repeated differentiation of a product of two functions. A simple relation between $A[f^{(n)}(u)](t)$ and $A[f(u)](t)$ follows directly from our definition. We have

$$A[f'(u)](r, \Delta x) = \frac{(-1)^r f^{(r+1)}(1/\Delta x)}{r!\Delta x^{r+1}} = -(r+1)\Delta x A[f(u)](r+1, \Delta x).$$

Hence, *for $t>0$, $A[f'(u)](t)$ and $A[f(u)](t)$ coexist, and satisfy the relation*

$$(36)\qquad A[f'(u)](t) = -tA[f(u)](t).^*$$

By induction, we obtain

$$(37)\qquad A[f^{(n)}(u)](t) = (-t)^n A[f(u)](t).$$

It is easily proved by the use of Cauchy's integral formula with circular contour for $f^{(n)}(z)$, and the corresponding conditions on $f(z)$, that, if $f(z)$ is of type zero over a certain sector, $f^{(n)}(z)$ is also of type zero over any sector interior to the sector of $f(z)$, and of the same angle. Likewise it can be shown that if $f(z)$ is of finite order $\rho$, $f^{(n)}(z)$ is of order $\rho-n$, or zero (according as $\rho>n$, or $\leqq n$), over any sector interior to that of $f(z)$, and of angle less than that of the sector of $f(z)$. However, the finite difference relation given above immediately shows that if $f(D)$ is of generalized order $\rho$, $f'(D)$ is of

* Comparison of this formula with (23) suggests a duality which is strikingly borne out, under special hypotheses, in a study of $A[f](t)$ for complex $t$, coupled with the extension to type one suggested in §17.

generalized order $\rho-1$ (or zero if $\rho=0$), with a corresponding extension to $f^{(n)}(D)$. A more complete statement is that $D^n f^{(n)}(D)$ is of the same order or generalized order as $f(D)$.

Turning now to the theorem desired, we shall first prove the special case

$$\{f(D)\}_{x_0}^{X} x\phi(x) = X\{f(D)\}_{x_0}^{X}\phi(x) + \{f'(D)\}_{x_0}^{X}\phi(x), \tag{38}$$

where $\phi(x)$ satisfies the conditions associated in our theorems with the existence of $\{f(D)\}_{x_0}^{X}\phi(x)$. We have, by definition,

$$\begin{aligned}\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X} x\phi(x) &= \sum_{r=0}^{[(X-x_0)/\Delta x]} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r!\Delta x^r}(X - r\Delta x)\phi(X - r\Delta x)\\ &= X\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x) + \left\{f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X-\Delta x}\phi(x)\\ &= X\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x) + \left\{f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x)\\ &\quad - \Delta x\left\{\frac{\Delta}{\Delta x}f'\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x).\end{aligned}$$

Since $Df'(D)$ is of the same type, or finite order, as $f(D)$, $\{Df'(D)\}_{x_0}^{X}\phi(x)$ will exist, so that the last term vanishes with $\Delta x$. Hence (38).

We can rewrite (38) in the form

$$\{f(D)\}_{x_0}^{X}(x - X)\phi(x) = \{f'(D)\}_{x_0}^{X}\phi(x),$$

and obtain, by induction,

$$\{f(D)\}_{x_0}^{X}(x - X)^n\phi(x) = \{f^{(n)}(D)\}_{x_0}^{X}\phi(x). \tag{39}$$

If then $P(x)$ is a polynomial of degree $m$, we can expand it in powers of $(x-X)$, and obtain, by (39), the terminating Leibnitz expansion

$$\begin{aligned}\{f(D)\}_{x_0}^{X}P(x)\phi(x) = P(X)\{f(D)\}_{x_0}^{X}\phi(x) &+ \frac{P'(X)}{1!}\{f'(D)\}_{x_0}^{X}\phi(x)\\ &+ \cdots + \frac{P^{(m)}(X)}{m!}\{f^{(m)}(D)\}_{x_0}^{X}\phi(x).\end{aligned}$$

This method will now be extended to $\{f(D)\}_{x_0}^{X}\psi(x)\phi(x)$.

We shall assume $\psi(x)$ to be analytic in the closed interval $(x_0, X)$ with the radius of convergence of its Taylor expansion at $x=X$ greater than $(X-x_0)$. Then, for this interval, we have

$$\psi(x) = \psi(X) + \frac{\psi'(X)}{1!}(x - X) + \frac{\psi''(X)}{2!}(x - X)^2 + \cdots.$$

When $f(D)$ is of finite order $\rho$, $f^{(\rho)}(D)$ is of order zero. Writing the remainder after $\rho$ terms of the $\psi(x)$ expansion $(x-X)^{\rho}\chi(x)$ we have, by (39),

$$\{f(D)\}_{x_0}^{X}(x - X)^{\rho}\chi(x)\phi(x) = \{f^{(\rho)}(D)\}_{x_0}^{X}\chi(x)\phi(x).$$

Since $f^{(\rho)}(D)$ is of order zero, the integral formula (15) is applicable. Furthermore, as the series for $\chi(x)$ is uniformly convergent over $(x_0, X)$, while the integral for $\{f^{(\rho)}(D)\}_{x_0}^{X}\phi(x)$ is absolutely convergent, we can integrate term by term. This is the same as operating with $f^{(\rho)}(D)$ term by term, or with $f(D)$ in the original series. We thus obtain, by (39), the generalized Leibnitz expansion

$$\{f(D)\}_{x_0}^{X}\psi(x)\phi(x) = \sum_{n=0}^{\infty}\frac{\psi^{(n)}(X)}{n!}\{f^{(n)}(D)\}_{x_0}^{X}\phi(x). \tag{40}$$

For $f(D)$ of type zero, we will designate the remainder after $m$ terms of the $\psi(x)$ series by $R_m(x)$, and prove directly that

$$\lim_{m\to\infty}\{f(D)\}_{x_0}^{X}R_m(x)\phi(x) = 0.$$

Applying Theorem II to this expression, we observe that $R_m(x)$ uniformly approaches zero as limit along $(x_0, x_1)$ as $m\to\infty$, so that the integral likewise has zero for limit. On the other hand, for $x_1$ sufficiently near $X$, the expansion of Theorem VI is applicable to $\{f(D)\}_{x_1}^{X}R_m(x)\phi(x)$. The first $N$ terms can be directly treated. On the other hand, the $h(N)$ of part (c) of §6, with $R_m(x)\phi(x)$ in place of $\phi(x)$, is an upper bound for the absolute value of the rest of this expansion. We thus see that

$$\left|\{f(D)\}_{x_0}^{X}R_m(x)\phi(x)\right| \leqq M_m L_m,$$

where $M_m$ is the upper bound of the function of the complex variable $R_m(z)\phi(z)$ over the circle center $x_1$, radius $l$, while $L_m$ depends only on $f(u)$, $(X-x_1)$, and the radius of convergence of the expansion of $R_m(z)\phi(z)$ about $z=x_1$. Now the radius of convergence of $R_m(z)$ is the same as that of $\psi(z)$, so that $L_m$ is independent of $m$. On the other hand, $R_m(z)$ uniformly approaches zero over the circle of radius $l$, so that $M_m\to 0$, as $m\to\infty$. The limit in question is thus proved to be zero, and we can state

THEOREM XII. *If $\{f(D)\}_{x_0}^{X}\phi(x)$ exists under the hypotheses of any of our existence theorems, and if $\psi(x)$ is analytic along the closed interval $(x_0, X)$, with the radius of convergence of its Taylor expansion at $x=X$ greater than $(X-x_0)$, then $\{f(D)\}_{x_0}^{X}\psi(x)\phi(x)$ is given by the generalized Leibnitz expansion* (40).

As a corollary, we have, by letting $\phi(x)$ identically equal 1,

$$\{f(D)\}^X_{x_0}\psi(x) = \psi(X)\{f(D)\}^X_{x_0}1 + \frac{\psi'(X)}{1!}\{f'(D)\}^X_{x_0}1 + \cdots, \tag{41}$$

which is to be compared with series (21), which may be written

$$\{f(D)\}^X_{x_0}\psi(x) = \psi(x_0)\{f(D)\}^X_{x_0}1 + \psi'(x_0)\{D^{-1}f(D)\}^X_{x_0}1 + \cdots. \tag{42}$$

In each case $\psi(x)$ is analytic along the closed interval $(x_0, X)$; but in the first the radius of convergence at $x=X$ exceeds $X-x_0$, in the second that at $x=x_0$ exceeds $X-x_0$.

We reserve for §14 the application of (41) to entire operators.

10. $e^{ax}\phi(x)$ as operand; $f(D+a)$. From our definition we have

$$A[f(u+a)](r, \Delta x) = \frac{(-1)^r f^{(r)}\left(\frac{1}{\Delta x} + a\right)}{r!\Delta x^{r+1}}.$$

For real $a$, and sufficiently small $\Delta x$, the equation

$$\frac{1}{\Delta x} + a = \frac{1}{\Delta_1 x}$$

results in a positive $\Delta_1 x$. We can then write

$$A[f(u+a)](r, \Delta x) = (1 - a\Delta_1 x)^{r+1} A[f(u)](r, \Delta_1 x).$$

As $\Delta x \to +0$, and $r\Delta x \to t$, we have simultaneously $\Delta_1 x \to +0$, $r\Delta_1 x \to t$. Hence under the sole condition that $a$ is real, we obtain

$$A[f(u+a)](t) = e^{-at}A[f(u)](t). \tag{43}$$

When $f(z)$ is of type zero, and hence also when it is of finite order, this result is obtained immediately, for $a$ both real and complex, by the use of the contour integral formula (8) of Theorem I.

The theorem of this section will be restricted to $f(D)$ of finite order. When $f(D)$ is of order zero, $f(D+a)$ also is. By (15), we have

$$\begin{aligned}\{f(D)\}^X_{x_0}e^{ax}\phi(x) &= \int_{x_0}^X A[f(u)](X-x)e^{ax}\phi(x)dx \\ &= e^{aX}\int_{x_0}^X e^{-a(X-x)}A[f(u)](X-x)\phi(x)dx,\end{aligned}$$

and so, by (43), and (15) again, we obtain

$$\{f(D)\}^X_{x_0}e^{ax}\phi(x) = e^{aX}\{f(D+a)\}^X_{x_0}\phi(x). \tag{44}$$

When $f(D)$ is of order $\rho$, $D^{-\rho}f(D)$ is of order zero. We then have

$$\begin{aligned}\{f(D)\}_{x_0}^{X}e^{az}\phi(x) &= d^{\rho}/dX^{\rho}\{D^{-\rho}f(D)\}_{x_0}^{X}e^{az}\phi(x)\\ &= e^{aX}(d/dX+a)^{\rho}\{(D+a)^{-\rho}f(D+a)\}_{x_0}^{X}\phi(x)\\ &= e^{aX}\{f(D+a)\}_{x_0}^{X}\phi(x);\end{aligned}$$

the first by (26), the second by (44) for operators of order zero and polynomials in $d/dX$, the last by (34). We can therefore state

THEOREM XIII. *If $f(D)$ is of finite order, and $\phi(x)$ satisfies the hypothesis of the existence Theorem* IV, *then relation* (44) *is valid.*

By letting $\phi(x)\equiv 1$, (44) and (8) yield a contour integral for $\{f(D)\}_{x_0}^{X}e^{ax}$.

11. $x_0=-\infty$; $\phi(x)=e^{ax}$. Let $f(z)$ be analytic to the right of the line $R(z)=c$. Then

$$\begin{aligned}\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^{X}e^{ax} &= f\left(\frac{1}{\Delta x}\right)e^{aX}-\frac{f'\left(\frac{1}{\Delta x}\right)}{1!\Delta x}e^{a(X-\Delta x)}+\cdots\\ &= e^{aX}f\left(\frac{1}{\Delta x}-\frac{e^{-a\Delta x}}{\Delta x}\right),\end{aligned}$$

provided $f(1/\Delta x-e^{-a\Delta x}/\Delta x)$ can be expanded in powers of $e^{-a\Delta x}/\Delta x$. Now

$$|e^{-a\Delta x}/\Delta x| = e^{-R(a)\Delta x}/\Delta x = 1/\Delta x - R(a) + \epsilon_{\Delta x},\quad \lim_{\Delta x\to+0}\epsilon_{\Delta x}=0.$$

Hence if $R(a)>c$, this absolute value will be less than $1/\Delta x-c$, for $\Delta x$ sufficiently small. As the radius of convergence of the Taylor expansion of $f(z)$ about $1/\Delta x$ is at least $1/\Delta x-c$, the above is valid, so that

$$\{f(D)\}_{-\infty}^{X}e^{ax} = \lim_{\Delta x\to+0}e^{aX}f\left(\frac{1-e^{-a\Delta x}}{\Delta x}\right) = e^{aX}f(a).$$

Conversely, if $f(z)$ is an analytic function of $z$, and $\{f(D)\}_{-\infty}^{X}e^{ax}$ exists for some $a$, $\{f(\Delta/\Delta x)\}_{-\infty}^{X}e^{ax}$ must converge for all positive $\Delta x$'s less than some positive $\epsilon$. Then $f(z)$ will be analytic within all circles center $1/\Delta x$, radius $|e^{-a\Delta x}/\Delta x|$ where $0<\Delta x<\epsilon$. As these circular regions cover the half-plane $R(z)>R(a)$, $f(z)$ is analytic to the right of the line $R(z)=R(a)$. Hence

THEOREM XIV. *The necessary and sufficient condition that $\{f(D)\}_{-\infty}^{X}e^{ax}$ exist for some $a$, when $f(z)$ is an analytic function of $z$, is that $f(z)$ is analytic to the right of some line $R(z)=c$. In that case $\{f(D)\}_{-\infty}^{X}e^{ax}$ exists for $R(a)>c$, and is given by*

$$\{f(D)\}_{-\infty}^{X}e^{ax} = e^{aX}f(a). \tag{45}$$

If $c<0$, we can let $a=0$, and so obtain

$$\{f(D)\}_{-\infty}^{X}1 = f(0). \tag{46}$$

The above results were obtained without assuming $f(z)$ of type zero, though the analyticity condition of Theorem XIV is a weaker form of the analyticity condition for an operator of type zero. If $f(z)$ is of type zero, we can apply Theorem XV of the next section to obtain

$$\lim_{x_0\to-\infty}\{f(D)\}_{x_0}^{X}e^{ax} = \{f(D)\}_{-\infty}^{X}e^{ax}. \tag{47}$$

This can also be obtained directly by expressing the remainder after $p$ terms of the Taylor expansion $\{f(\Delta/\Delta x)\}_{-\infty}^{X}e^{ax}$ as a contour integral, finding for its limit as $\Delta x\to+0$, $p=\{(X-x_0)/\Delta x\}$, an integral like that of formula (8), and observing that this approaches zero as limit as $x_0\to-\infty$. If we further restrict $f(z)$ to order zero, we can use (15) in (47), and by the change of variable $X-x=t$, obtain through (45)

$$\int_0^\infty e^{-at}A[f](t)dt = f(a).* \tag{48}$$

That is, $A[f](t)$ satisfies the Laplace integral equation when $f(z)$ is of order zero. The derivation of (48) suggests that we may consider relation (45) a generalization of the Laplace integral equation.

12. $x_0=-\infty$: **general case.** Our definition, for $x_0=-\infty$, is

$$\{f(D)\}_{-\infty}^{X}\phi(x) = \lim_{\Delta x\to+0}\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^{X}\phi(x).$$

Hence the relation

$$\{f(D)\}_{-\infty}^{X}\phi(x) = \lim_{x_0\to-\infty}\{f(D)\}_{x_0}^{X}\phi(x) \tag{49}$$

is subject to proof. Since, in our definition, we consider the limit of an infinite series, some restriction on the behavior of $\phi(x)$ as $x\to-\infty$ will have to be introduced to insure the existence of the limits involved. The same condition will turn out to be adequate for (49).

We assume $f(D)$ to be of type zero. Then $f(z)$ will certainly be analytic

* When $A[f](t)$ can be directly found from its definition, formula (48) leads to the evaluation of a corresponding definite integral. Thus, (6), the formula for $A[u^n](t)$, yields the well known infinite integral for the gamma function for $R(n)<0$, while (3), as transformed by (12), results in an infinite integral for the Eulerian constant $\gamma$. In a similar manner, formula (8) leads to the evaluation of corresponding contour integrals.

to the right of some line $R(z)=c$. *Let $d$ designate the lower limit of these $c$'s.* Then, *for $\Delta x$ sufficiently small, and $r\Delta x$ greater than a positive $h$, there is a positive $L$ for each $b$ greater than $d$ such that*

$$|A[f](r,\Delta x)| < Le^{b(r\Delta x)}. \tag{50}$$

To prove this fundamental inequality, we reconsider the discussion of the contour integral for $A[f](r,\Delta x)$ given in §3. For $\Delta x$ sufficiently small, and $r\Delta x>h$, $r$ will be sufficiently large to have, along $C''$,

$$\left|\int_{C''}\frac{f(z)dz}{(1-z\Delta x)^{r+1}}\right| < L_1\lambda^r,\ 0<\lambda<1.$$

Since, for $\Delta x$ sufficiently small, we shall have $\lambda$ less than $e^{b\Delta x}$ even with $b$ negative, a relation like (50) holds for this contribution to $A[f](r,\Delta x)$. Along $C'_{l,m}$, we can have, with $l$ and $m$ sufficiently large, $R(z)<b'$, where $b'<b$, and $b'<0$. We can then write

$$\left|\int_{C'_{l,m}}\frac{f(z)dz}{(1-z\Delta x)^{r+1}}\right| < \frac{1}{(1-b'\Delta x)^{r+1-[h/\Delta x]}}\int_{C'_{l,m}}\frac{|f(z)||dz|}{|1-z\Delta x|^{[h/\Delta q]}}.$$

The latter integral has a finite upper bound, for $\Delta x$ sufficiently small, in accordance with §3. On the other hand, since $<b$, we have, again for sufficiently small $\Delta x$,

$$\frac{1}{1-b'\Delta x} < e^{b\Delta x}.$$

Hence an inequality (50) exists for the $C'_{l,m}$ contribution. Finally, we may have to change $C_{l,m}$ so that it will consist of the segment of $R(z)=b''$, $d<b''<b$, cut off by the former $C_{l,m}$, joined to the part of the former $C_{l,m}$ to the left of $R(z)=b''$. Since along this new $C_{l,m}$ we have

$$|1-z\Delta x| \geqq 1-b''\Delta x,$$

we find that

$$\left|\int_{C_{l,m}}\frac{f(z)dz}{(1-z\Delta x)^{r+1}}\right| \leqq \frac{\bar{l}K'}{(1-b''\Delta x)^{r+1}},$$

where $\bar{l}$ is the length of $C_{l,m}$, and $K'$ the upper bound of $|f(z)|$ along it. Hence, as in the case of $C'_{l,m}$, a relation (50) holds. By combining these three results we get (50) itself.

If in (50) we let $\Delta x\to+0$, $r\Delta x\to t$, we obtain, for $t>h$,

$$|A[f](t)| \leqq Le^{bt}. \tag{51}$$

We can now prove

THEOREM XV. *If $f(D)$ is of type zero, with specified $d$, and if, for some positive $M$, and real $c$ greater than $d$, $\phi(x)$ satisfies the inequality*

$$|\phi(x)| \leqq Me^{cx}$$

*for every $x$ less than some real $x_1$, relation* (49) *will be valid.*

Choose $b$ between $c$ and $d$. By (50), and the condition on $\phi(x)$, we have

$$\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^{X}\phi(x) - \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x)\right|$$

$$< LMe^{cX}\sum_{r=p+1}^{\infty}e^{-(c-b)r\Delta x}\Delta x, \qquad p = \left\{\frac{X - x_0}{\Delta x}\right\}$$

provided $x_0 < x_1$, and also $x_0 < X - h$. Since this geometric progression converges, $\{f(\Delta/\Delta x)\}_{-\infty}^{X}\phi(x)$ also converges. Furthermore, on summing this progression, we find its limit, as $\Delta x \to +0$, to be $[LM/(c-b)]e^{cX}e^{-(c-b)(X-x_0)}$. This expression approaches zero as limit as $x_0 \to -\infty$. We can therefore apply the limit criterion, with $x_0$ in place of $\nu$, to the identity

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^{X}\phi(x) = \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x)$$

$$+\left[\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{-\infty}^{X}\phi(x) - \left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{x_0}^{X}\phi(x)\right],$$

and obtain the theorem.

The use of the limit criterion assumes the existence of $\{f(D)\}_{x_0}^{X}\phi(x)$ for all $x_0$'s less than $X$. Granting this, its corollary gives the existence of both members of (49). Hence

COROLLARY I. *If $\phi(x)$ is continuous for $x < X$, and satisfies the $X$-neighborhood conditions of any of our existence theorems for finite $x_0$, then, under the above hypothesis, both members of* (49) *exist, and are finite.*

It is also not difficult to establish

COROLLARY II. *If the $X$-neighborhood condition of Corollary* I *is not known to be fulfilled, but instead $\{f(D)\}_{-\infty}^{X}\phi(x)$ is known to exist, then the existence of $\{f(D)\}_{x_0}^{X}\phi(x)$ for finite $x_0$ follows.*

The rest of this section will be devoted to the law of successive operations for $x_0 = -\infty$. For finite $x_0$, the order of $f_2(D)$ had to be zero or one, unless $\phi(x)$, and a sufficient number of its derivatives, vanished at $x = x_0$. In the present case this restriction on the order of $f_2(D)$ disappears. We shall, however, still restrict ourselves to operators of finite order.

First, in Theorem XV, let $f(D)$ be of order zero. Then, by (49) and (15), we obtain

$$\{f(D)\}_{-\infty}^{X}\phi(x) = \int_{-\infty}^{X} A[f](X-x)\phi(x)dx. \tag{52}$$

Break up this integral over limits $(-\infty, X-a)$, and $(X-a, X)$. By applying to the resulting integrals inequality (51), with $b$ between $c$ and $d$, and the hypothesis of Theorem XV, we obtain, for $X<x_1$, and $a>h$,

$$\left|\int_{-\infty}^{X} A[f](X-x)\phi(x)dx\right| < Me^{cX+|c|a}\int_{0}^{a} |A[f](t)|\,dt + [LM/(c-b)]e^{-a(c-b)}\cdot e^{cX} = M'e^{cX}.$$

That is, $|\{f(D)\}_{-\infty}^{X}\phi(x)|$ satisfies the same inequality that $|\phi(x)|$ satisfies except for the factor $M$. It is also easily seen that $\{f(D)\}_{-\infty}^{X}\phi(x)$ is continuous in $X$.

Let then $f_1(D)$ and $f_2(D)$ both be of order zero, with $d$'s equal to $d_1$ and $d_2$ respectively. If, in the hypothesis of Theorem XV, we have $c>d_1$ as well as $c>d_2$, our last result shows that the existent $\{f_2(D)\}_{-\infty}^{X}\phi(x)$ can be used as operand for $\{f_1(D)\}_{-\infty}^{X}$, so that we can write

$$\{f_1(D)\}_{-\infty}^{X}\{f_2(D)\}_{-\infty}^{x}\phi(\xi) = \int_{-\infty}^{X} A[f_1](X-x)\left[\int_{-\infty}^{x} A[f_2](x-\xi)\phi(\xi)d\xi\right]dx.$$

Since this iterated integral is absolutely convergent, the corresponding improper double integral exists.* It can therefore be rewritten

$$\int_{-\infty}^{X}\left[\int_{x}^{X} A[f_1](X-\xi)A[f_2](\xi-x)d\xi\right]\phi(x)dx.$$

Now the $d$ of $f_1(D)f_2(D)$ does not exceed both $d_1$ and $d_2$. We can therefore use (52), with $f_1(D)f_2(D)$ as operator, $\phi(x)$ as operand, and, by applying (32), obtain

$$\{f_1(D)\}_{-\infty}^{X}\{f_2(D)\}_{-\infty}^{x}\phi(\xi) = \{f_1(D)f_2(D)\}_{-\infty}^{X}\phi(x). \tag{53}$$

Before (53) can be extended to operators of arbitrary finite order, several preliminary results must be obtained. First, we have

$$\frac{d}{dX}\{f(D)\}_{-\infty}^{X}\phi(x) = \{Df(D)\}_{-\infty}^{X}\phi(x) \tag{54}$$

under the same $X$-neighborhood conditions as for finite $x_0$, along with the

* The reasoning is the same as in §8 for which a reference has been given.

usual inequality on $|\phi(x)|$.* In fact, our inequalities easily show that the integral resulting from differentiating $\int_{-\infty}^{x_2} A[f](X-x)\phi(x)dx$ under the integral sign is uniformly convergent in $X$. Secondly, we have to discard the reduction formula (17), since the $d$ of $D^{-\rho}f(D)$ will be greater than that of $f(D)$ when the latter is negative and $\rho$ is not zero. By §10 and formula (17) with $x_1$ and $x_0$ identical, we obtain in its place

$$\{f(D)\}_{x_0}^{X}\phi(x) = \{(D-l)^{-\rho}f(D)\}_{x_0}^{X}\left(\frac{d}{dx}-l\right)^{\rho}\phi(x) + \sum_{\mu=0}^{\rho-1} A[(u-l)^{-\mu-1}f(u)](X-x_0)\cdot\left[\left(\frac{d}{dx}-l\right)^{\mu}\phi(x)\right]_{x=x_0}, \tag{55}$$

where, for validity, $\phi(x)$ and its first $\rho$ derivatives are continuous in $(x_0, X)$.

To allow $x_0$ to approach $-\infty$ as limit in (55), we shall want $\phi(x)$, and its first $\rho$ derivatives, to be continuous for $x \leqq X$. Also, for some $c$ greater than $d$, and some positive $M$, we must have, for $x<x_1$,

$$|\phi^{(\mu)}(x)| \leqq Me^{cx}; \quad \mu = 0, 1, 2, \cdots, \rho.\dagger$$

As a result, we also have, for $x<x_1$,

$$\left|\left(\frac{d}{dx}-l\right)^{\mu}\phi(x)\right| \leqq M'e^{cx}; \quad \mu = 0, 1, 2, \cdots, \rho.$$

Now choose $l$ less than $c$. The $d$ of $(D-l)^{-\mu-1}f(D)$ will then also be less than $c$. Choose $b$ between this $d$ and $c$. Then, by (51),

$$|A[(u-l)^{-\mu-1}f(u)](X-x_0)| \leqq L_\mu e^{b(X-x_0)}.$$

As $b<c$, these two inequalities give us

$$\lim_{x_0\to-\infty} A[(u-l)^{-\mu-1}f(u)](X-x_0)\left[\left(\frac{d}{dx}-l\right)^{\mu}\phi(x)\right]_{x=x_0} = 0.$$

We thus obtain

$$\{f(D)\}_{-\infty}^{X}\phi(x) = \{(D-l)^{-\rho}f(D)\}_{-\infty}^{X}\left(\frac{d}{dx}-l\right)\phi(x). \tag{56}$$

Though (56) reduces all operators of finite order to operators of order zero, for which we have already established the law of successive operations, its conditions are stronger than those yielded by the following more extended

* See the first footnote of §8.

† If $\phi(x)$ behaves "regularly" as $x\to-\infty$, the inequalities for $\mu>0$ will follow from the one for $\mu=0$.

treatment. The operators $f_1(D)$ and $f_2(D)$ will now be of arbitrary finite orders $\rho_1$ and $\rho_2$ respectively with corresponding $d_1$ and $d_2$. We shall require the inequalities

$$|\phi^{(\mu)}(x)| \leqq M_1 e^{c_1 x}, \quad \mu = 0, 1, 2, \cdots, \rho_2, \quad \text{for} \quad x < x_1, \quad \text{with} \quad c_1 > d_1.$$

When $d_2 \geqq c_1$, we shall also require, to insure the existence of $\{f_2(D)\}^x_{-\infty}\phi(\xi)$,

$$|\phi(x)| \leqq M_2 e^{c_2 x}, \text{ for } x < x_1, \text{ with } c_2 > d_2.$$

In the latter case $c_2 > c_1$, so that, for some $x_2$, we shall have for $x < x_2$, $M_2 e^{c_2 x} < M_1 e^{c_1 x}$. In either case, we can prove that, if $l$ is less than $c_1$,

$$\left| \frac{d^\mu}{dx^\mu} \{(D - l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi) \right| \leqq M_3 e^{c_1 x}, \text{ for } x < x_3, \ \mu = 0, 1, 2, \cdots, \rho_2.$$

In fact, through (54) extended, we can write

$$\frac{d^\mu}{dx^\mu} \{(D - l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi) = \{D^\mu (D - l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi)$$

$$+ \int_{-\infty}^{x-a} A\,[u^\mu (u - l)^{-\rho_2} f_2(u)\,](x - \xi)\phi(\xi)\,d\xi.$$

By reducing the first term of the right hand member of this equation by formula (17) with $\rho = \mu$, $X = x$, $x_1 = x_0 = x - a$, we can easily apply our inequalities, by methods already made familiar, to obtain the desired result. We can therefore use (56), with $\rho_2$ in place of $\rho$, $(D-l)^{-\rho_1} f_1(D)$ in place of $f(D)$, and $\{(D-l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi)$ in place of $\phi(x)$ to obtain

$$\{(D - l)^{-\rho_1} f_1(D)\}^X_{-\infty} \{(D - l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi)$$

$$= \{(D - l)^{-(\rho_1+\rho_2)} f_1(D)\}^X_{-\infty} \left(\frac{d}{dx} - l\right)^{\rho_2} \{(D - l)^{-\rho_2} f_2(D)\}^x_{-\infty}\phi(\xi).$$

Now reduce the first member of this equation by our proved result for operators of order zero, the second member by operating with $(d/dx - l)^{\rho_2}$ formally, as is justified by (54). We thus obtain

$$\{(D - l)^{-(\rho_1+\rho_2)} f_1(D) f_2(D)\}^X_{-\infty}\phi(x) = \{(D - l)^{-(\rho_1+\rho_2)} f_1(D)\}^X_{-\infty} \{f_2(D)^x_{-\infty}\}\phi(\xi).$$

Operating on both sides of this equation by $(d/dx - l)^{\rho_1+\rho_2}$ yields our law of successive operations for operators of arbitrary finite orders.

In this discussion we have merely concerned ourselves with the convergence difficulties introduced by letting $x_0 = -\infty$. When we also supply the discussion of the existence of derivatives tacitly assumed above, a discussion that does not introduce anything essentially new, we obtain

THEOREM XVI. *Let $f_1(D)$ and $f_2(D)$ be of finite orders $\rho_1$ and $\rho_2$ respectively, with corresponding $d_1$ and $d_2$, and let $D^{\rho}f_2(D)$ be of order $\rho$. Let $\phi(x)$ and its first $\rho_2$ derivatives be continuous for $x \leqq X$, and also let the first $\rho$ derivatives of $\phi(x)$ be continuous in a left neighborhood of $x=X$. Finally, let us have the inequalities*

$$|\phi^{(\mu)}(x)| \leqq M_1 e^{c_1 x}, \text{ for } x < X, \ \mu = 0, 1, 2, \cdots, \rho_2, \text{ with } c_1 > d_1,$$

*and, if $c_1$ does not exceed $d_2$, let us also have*

$$|\phi(x)| \leqq M_2 e^{c_2 x}, \text{ for } x < X, \text{ with } c_2 > d_2.$$

*Then formula* (53) *is valid, and both of its members exist.*

13. $x_0 = -\infty$: **special case.** When $d=0$, the existence theorem of the preceding section still requires an exponential inequality for $|\phi(x)|$ as $x \to -\infty$. This can be replaced by an algebraic inequality if certain assumptions are made about the behavior of $f(z)$ in the neighborhood of those singularities that lie on the axis of imaginaries. The operator $D^m$ is typical.

Let then $f(z)$ be of type zero, and analytic to the right of the axis of imaginaries. We shall first assume the origin to be the only singularity with real part zero. The assumptions required are the following:

(a) *$f(z)$ is analytic within a circular sector of radius $\lambda$, and angle greater than $\pi$, which has its vertex at the origin, and is bisected by the positive half of the real axis,*

(b) *in this sector, for some real $m$ and positive $K$, $|f(z)|$ satisfies the inequality*

$$|f(z)| \leqq K|z|^m.$$

Let $S$ designate the interior of the infinite analytic sector of the type zero condition, $S'$ the rest of the $z$-plane; $s$ the interior of the circular sector of radius $\lambda$, $s'$ the rest of the interior of the circle. It will be convenient to have $\lambda$ small enough for this circle to lie wholly within $S'$. Since the origin is the only singularity with real value zero, there will be a line $R(z) = -\kappa_1$, $\kappa_1 > 0$, to the right of which $f(z)$ is analytic, except for $z$'s in $s'$. Now in the argument leading to inequality (50) of the preceding section, replace $d$ by $-\kappa_1$, and choose the $b''$ of that argument between 0 and $-\kappa_1$, and also sufficiently near zero to have the line $R(z)=b''$ cut the sides of $s'$. If then we remove from $C'''$, as modified in that argument, the portion of this line that is in $s'$, we shall have along the resulting open contour $C^{\text{iv}}$

$$\left| \int_{C^{\text{iv}}} \frac{f(z)dz}{(1 - z\Delta x)^{r+1}} \right| < L_1 e^{-\kappa(r\Delta x)}$$

where $-\kappa$ is between 0 and $b''$, and hence is negative.* For the purposes of this section we further modify $C'''$ as follows. Let $C_\delta^{\text{v}}$ consist of two segments in $s$, starting from a point $z=\delta$ on the real axis, running parallel to the sides of $s$, and terminated by $R(z)=b''$. Such a contour is possible for $\delta$ sufficiently small. Join this to $C^{\text{iv}}$, where we remove from $C^{\text{iv}}$ the portions of $R(z)=b''$ that lie between the ends of $C_\delta^{\text{v}}$. Then $C_\delta^{\text{v}}$ with the shortened $C^{\text{iv}}$ forms an admissible $C'''$. Now if $C^{\text{iv}}$ is thus shortened in the above inequality, the result is simply strengthened. As for $C_\delta^{\text{v}}$, replace $\delta$ by $1/[2(r+1)\Delta x]$, and transform $z$ by $\zeta=(r+1)\Delta x z$. Using condition (b), we obtain

$$\left|\int_{C_\delta^{\text{v}}} \frac{f(z)dz}{(1-z\Delta x)^{r+1}}\right| \leqq \frac{K}{[(r+1)\Delta x]^{m+1}} \int_{C_{1/2}^{\text{v}}} \frac{|\zeta|^m |d\zeta|}{\left|1-\frac{\zeta}{r+1}\right|^{r+1}}.$$

If the sides of $C_{1/2}^{\text{v}}$ be extended to infinity, we obtain the same integral that occurred in §11, except that $m$ replaces $\rho-\epsilon$. This integral was there shown to converge, and be bounded with respect to $r$ for $r \geqq \rho$. The same is thus true of the $C_{1/2}^{\text{v}}$ integral for $r>m$. Combining these results, we are enabled to state that, *for $\Delta x$ sufficiently small, and $r\Delta x$ greater than a fixed positive $h$, we have with a positive $L$,*

$$|A[f](r,\Delta x)| < \frac{L}{(r\Delta x)^{m+1}}. \tag{57}$$

Hence, also, for $t$ greater than $h$, we have

$$|A[f](t)| \leqq \frac{L}{t^{m+1}}. \dagger \tag{58}$$

This result can be immediately extended to the case where there are a finite number of singularities of the above type on the axis of imaginaries. If these singularities are at points $z=z_i$, then on each we impose conditions like (a) and (b), where in (a) the vertex of the sector is now at $z_i$, while the inequality in (b) is replaced by

$$|f(z)| \leqq K_i |z-z_i|^{m_i}.$$

The line $R(z)=-\kappa_1$ can be chosen once for all, and $C'''$ modified near each of the singularities. As a result, (57) and (58) will follow, provided *$m$ is*

* Inequality (b) of this section, in conjunction with the choice of $b''$ made here, allows us to conclude that $|f(z)|$ has a finite upper bound along that part of $R(z)=b''$ used in $C^{\text{iv}}$, as is required in this proof.

† This inequality can be used as the basis of a simple derivation of certain of the Heaviside asymptotic expansions, and extensions thereof.

*the least of the $m_i$'s.* For ease of reference we shall speak of a function and corresponding operator of this kind as being of *degree m.* We can now state

THEOREM XVII. *If an operator $f(D)$ of type zero, with $d=0$, is of degree $m$, and if for $x<x_1$, $x_1<0$, we have*

$$|\phi(x)| \leqq M(-x)^{m-\eta}$$

*with positive $\eta$ and $M$, the equivalence relation* (49) *will hold.*

COROLLARY. *With this hypothesis replacing that of Theorem* XV, *Corollary* I *and Corollary* II *of the latter theorem continue to hold here.*

In the proof of this theorem the geometric progression of the preceding section is replaced by a series that can be written

$$LM\sum_{r=p+1}^{\infty}\left(1-\frac{X}{r\Delta x}\right)^{m+1}\frac{1}{(r\Delta x-X)^{1+\eta}}\Delta x,$$

provided $x_0<x_1$, and, also, $x_0<X-h$. Since $[1-X/(r\Delta x)]^{m+1}$ is bounded as $\Delta x\to+0$, and $x_0\to-\infty$, while we have

$$\sum_{r=p+1}^{\infty}\frac{1}{(r\Delta x-X)^{1+\eta}}\Delta x<\int_{X-x_0-\Delta x}^{\infty}(t-X)^{-1-\eta}dt=\frac{(-x_0-\Delta x)^{-\eta}}{\eta},$$

recourse can again be had to the limit criterion to give the theorem.

The case $\phi(x)\equiv 1$ is of special interest. When $d<0$, the discussion of §11 applies, as well as the resulting formula (46). When $d>0$, the same discussion shows $\{f(D)\}_{-\infty}^{X}1$ to be non-existent. When $d=0$, (46) still holds, provided $f(z)$ is analytic at the origin; but the equivalence relation (49) seems to require further assumptions. The following theorem, which we state without proof, contains several simpler cases, and can be considered the extension of (46) for $d=0$.

THEOREM XVIII. *If $f(z)$ is of type zero, with $d=0$, and if, except for the origin, it is of degree of greater than $-1$, while for the analytic sector vertexed at the origin it approaches a finite limit $K$ as $z\to 0$ in the sector, then*

$$\{f(D)\}_{-\infty}^{X}1=K=\lim_{x_0\to-\infty}\{f(D)\}_{x_0}^{X}1. \tag{59}$$

We turn now to the law of successive operations. As before, $f_1(z)$ and $f_2(z)$ will be of finite order. The case of most interest is the one in which both $d_1$ and $d_2$ are zero, the treatment of the cases where but one $d$ is zero then being obvious. On the whole, the development of the preceding section can be followed.

First let $f(D)$ be of order zero, while otherwise it, and $\phi(x)$, satisfy the hypothesis of Theorem XVII. We can again use (46), and break up the infinite integral over limits $(-\infty, X-a)$, $(X-a, X)$. Choose $a$ greater than $h$, and let $X$ be less than $x_1$. We then easily obtain

$$\left|\int_{X-a}^{X} A[f](X-x)\phi(x)dx\right| \leqq M(-X+\theta a)^{m-\eta}\int_0^a |A[f](t)|\,dt < M''(-X)^{m-\eta},$$

$$\left|\int_{-\infty}^{X-a} A[f](X-x)\phi(x)dx\right| < LM\int_{-\infty}^{X-a}\frac{(-x)^{m-\eta}}{(X-x)^{m+1}}dx$$

$$= LM(-X)^{-\eta}\int_0^{-X/a}(1+t)^{m-\eta}t^{\eta-1}dt,$$

where, in the latter, we have set $t=-X/(X-x)$. It is convenient to assume $X<-a$, so that $-X/a>1$. We can then break up this integral over limits $(0,1)$, and $(1,-X/a)$. The first part converges, and is independent of $X$. For the second, we can write, with some fixed $k$,

$$\int_1^{-X/a}(1+t)^{m-\eta}t^{\eta-1}dt < k\int_1^{-X/a}t^{m-1}dt.$$

The case $m=0$ can be avoided. For $m\neq 0$ we thus obtain the inequalities

$$(60)\qquad m>0: \quad |\{f(D)\}_{-\infty}^{X}\phi(x)| < M'(-X)^{m-\eta},$$

$$(61)\qquad m<0: \quad |\{f(D)\}_{-\infty}^{X}\phi(x)| < M'(-X)^{-\eta}.$$

Now let $f_1(D)$ and $f_2(D)$ both be of order zero, with $d_1$ and $d_2$ zero, and associated $m_1$ and $m_2$. No loss of generality ensues if we assume the $d$ of $f_1(D)f_2(D)$ to be zero.* The corresponding $m_3$ is evidently not less than $m$, the least of the three quantities $m_1$, $m_2$, $m_1+m_2$. We can then show that if $\phi(x)$ satisfies the hypothesis of Theorem XVII for $m$ so defined, the law of successive operations will be satisfied. It will be sufficient to show that $\{f_2(D)\}_{-\infty}^{x}\phi(\xi)$, $\{f_1(D)\}_{-\infty}^{X}\{f_2(D)\}_{-\infty}^{x}\phi(\xi)$ and $\{f_1(D)f_2(D)\}_{-\infty}^{X}\phi(x)$ exist in accordance with Theorem XVII, since the argument of the preceding section will then be valid here. The inequality of Theorem XVII is the primary consideration. Since the three operators involved, $f_2(D)$, $f_1(D)$, $f_1(D)f_2(D)$, are of degrees $m_2$, $m_1$, and $m_3$ respectively, we have to show that the corresponding operands $\phi(\xi)$, $\{f_2(D)\}_{-\infty}^{x}\phi(\xi)$ and $\phi(x)$ satisfy inequalities with exponents $m_2-\eta_2$, $m_1-\eta_1$ and $m_3-\eta_3$ respectively, where the $\eta$'s are all positive. Note that our hypothesis makes $\phi(x)$ satisfy an inequality with

* This $d$ may be less than zero; but in that case the actual inequalities are even stronger than on the assumption that it is zero.

exponent $m-\eta$ where $\eta$ is positive. This suffices for the first and last cases; for by identifying this inequality with the desired inequalities, we are setting $m_2-\eta_2=m-\eta$, $m_3-\eta_3=m-\eta$. Since $m$ does not exceed $m_2$ or $m_3$, we thus have $\eta_2 \geqq \eta$, $\eta_3 \geqq \eta$, that is, $\eta_2$ and $\eta_3$ are positive, as was desired. For $\eta_1$, we must use the inequalities (60) and (61). The case $m_2=0$ can be avoided since $m_2$ can always be decreased in the inequalities on $|f_2(z)|$, at least for $|z-z_i| \leqq 1$, and by making this decrease less than $\eta$ there will at worst result a corresponding decrease in $m$ and $\eta$ in the original inequality on $\phi(x)$, with the new $\eta$ still positive. For $m_2>0$, we see from (60) that $\{f_2(D)\}_{-\infty}^{x}\phi(\xi)$ satisfies an inequality with exponent equal to that of the inequality for $\phi(\xi)$. That is, we may set $m_1-\eta_1=m-\eta$, and as before, obtain $\eta_1 \geqq \eta$. For $m_2<0$, (61) shows that the exponent for $\{f_2(D)\}_{-\infty}^{x}\phi(\xi)$ is but $-\eta_2$, if $m_2-\eta_2$ is the exponent for $\phi(\xi)$. That is, we may set $m_1-\eta_1=-\eta_2$, where we have identically $m_2-\eta_2=m-\eta$. By combining the two equations, we obtain $\eta_1=m_1+m_2-m+\eta \geqq \eta$. A positive $\eta_1$ therefore results. The sufficiency of the $m-\eta$ exponent for $\phi(x)$ has thus been demonstrated.

It is unnecessary to go into the details of the extension of these results to operators of arbitrary finite orders, as the steps of the procedure of the previous section are easily verified here. We can thus state

THEOREM XIX. *Let $f_1(D)$ and $f_2(D)$ be of finite orders $\rho_1$ and $\rho_2$ respectiveiy with $d_1$ and $d_2$ zero, and associated $m_1$ and $m_2$, and let $D^{\rho}f_2(D)$ be of order $\rho$. Let $\phi(x)$ and its first $\rho_2$ derivatives be continuous for $x \leqq X$, and also let the first $\rho$ derivatives of $\phi(x)$ be continuous in a left neighborhood of $x=X$. Finally, for all $x$'s such that $x<x_1<0$, let*

$$|\phi(x)| \leqq M(-X)^{m-\eta}, \quad |\phi^{(\mu)}(x)| \leqq M_1(-X)^{m_1-\eta_1}, \quad \mu = 1, 2, \cdots, \rho_2,$$

*where $\eta$ and $\eta_1$ are positive, and $m$ is the least of the three quantities $m_1$, $m_2$, $m_1+m_2$. Then formula (53) holds, and both of its members exist.*

14. **Entire operators; $A[f](t) \equiv 0$.** If $f(z)$ is an entire function of the complex variable $z$, and satisfies condition (b) of §3 over the whole $z$-plane, then $f(D)$ will be said to be of *type zero over the plane*. Since entire transcendental functions of genus zero are known to satisfy this condition, their corresponding operators are of type zero over the plane. Polynomials in $D$ are also included.

We first prove that, for such operators, $A[f](t)$ *vanishes identically*. In formula (8) of §3 choose $C$ so that its vertex is at the origin. Let $\overline{C}_N$ be the portion of the line $R(z)=-N$, $N>0$, between the half-lines of $C$. Since $f(z)$ is entire, the integral in (8), limited to that part of $C$ which is to the right of this line, will equal the same integral over $\overline{C}_N$. Hence we have

$$A[f](t) = \frac{1}{2\pi i} \lim_{N\to\infty} \int_{\bar{C}_N} e^{tz} f(z)dz.$$

By choosing $\kappa$ of condition (b) less than $t \cos \alpha$, this limit is seen to be zero.

An immediate consequence of this result is that, for continuous $\phi(x)$, $\{f(D)\}_{x_0}^{X}\phi(x)$ *is independent of the finite lower limit* $x_0$, since by (9),

$$\{f(D)\}_{x_0}^{X}\phi(x) = \{f(D)\}_{x_1}^{X}\phi(x).$$

The same is evidently true for $x_0 = -\infty$, if $\phi(x)$ satisfies the hypothesis of Theorem XV. We can now easily show that when $f(D)$ is of type zero over the plane, $\{f(D)\}_{x_0}^{X}\phi(x)$ *reduces to the formal result obtained by expanding* $f(D)$ *in powers of* $D$. The function $\phi(x)$ is of course assumed to be analytic in a left neighborhood of $x = X$. We can then choose $x_1$ sufficiently near $X$ to apply the Leibnitz expansion (41) to $\{f(D)\}_{x_1}^{X}\phi(x)$. We thus have

$$\{f(D)\}_{x_0}^{X}\phi(x) = \phi(X)\{f(D)\}_{x_1}^{X}1 + \frac{\phi'(X)}{1!}\{f'(D)\}_{x_1}^{X}1 + \cdots.$$

Now $f^{(n)}(D)$ is of type zero over the plane for every $n$. Hence

$$\{f^{(n)}(D)\}_{x_1}^{X}1 = \{f^{(n)}(D)\}_{-\infty}^{X}1 = f^{(n)}(0),$$

the last by (46). We therefore have the standard expansion

$$\{f(D)\}_{x_0}^{X}\phi(x) = f(0)\phi(X) + \frac{f'(0)}{1!}\phi'(X) + \frac{f''(0)}{2!}\phi''(X) + \cdots. \tag{62}$$

Suppose now, conversely, that $\{f(D)\}_{x_0}\phi(x)$ is independent of $x_0$ in a certain $x_0$ interval. Differentiation of (9) with respect to $x_0$ gives

$$A[f](X - x)\phi(x_0) = 0.$$

If then $\phi(x_0)$ does not vanish in this interval, $A[f](t)$ must vanish over a corresponding interval. Now formula (8) shows $A[f](t)$ to be an analytic function of $t$ for every real and positive $t$. Hence if $A[f](t)$ vanishes over an interval, it vanishes identically. The question thus raised is answered by

THEOREM XX. *If* $f(D)$ *is of finite order,* $A[f](t)$ *vanishes identically when, and only when,* $f(D)$ *is a polynomial; if* $f(D)$ *is of type zero,* $A[f](t)$ *vanishes identicallly when, and only when,* $f(D)$ *is of type zero over the plane.*

The direct part of this theorem has already been demonstrated. As for the converses, let $f(D)$ first be of finite order $\rho$. Then $f^{(\rho)}(D)$ will be of order zero, so that, for $R(a) > d$, we shall have, by (48),

$$f^{(\rho)}(a) = \int_0^\infty e^{-at} A[f^{(\rho)}](t)dt.$$

Since under our hypothesis $A[f](t) \equiv 0$, (37) gives us $A[f^{(\rho)}](t) \equiv 0$, so that $f^{(\rho)}(a) = 0$ for all $a$'s in at least a half-plane. $f(a)$ is therefore a polynomial in that half-plane, and, being analytic, is a polynomial throughout.

Now let $f(D)$ be of type zero, with $A[f](t) \equiv 0$. Then, for $R(a) > d$,

$$\{f(D)\}_{-\infty}^{X} e^{ax} = \{f(D)\}_{x_1}^{X} e^{ax}.$$

We can apply (45) to the first member of this equation; and, as $e^{ax}$ is an entire function of $x$, (41) will yield a convergent series for the second. We thus get

$$f(a) = \{f(D)\}_{x_1}^{X} 1 + \frac{\{f'(D)\}_{x_1}^{X} 1}{1!} a + \frac{\{f''(D)\}_{x_1}^{X} 1}{2!} a^2 + \cdots.$$

Since this power series in $a$ converges in a half-plane of $a$, its radius of convergence must be infinite, and so $f(a)$ is entire. To prove that it is of type zero over the plane, note that, from the series, we have

$$\{f^{(n)}(D)\}_{x_1}^{X} 1 = f^{(n)}(0).$$

Now let $X$ be any positive number, however small, and choose $x_1$ between 0 and $X$. Formula (41) will then yield the convergent series

$$\{f(D)\}_{x_1}^{X} x^{-1} = X^{-1} f(0) - X^{-2} f'(0) + X^{-3} f''(0) - \cdots.$$

Since this converges for every positive $X$ we must have

$$\lim_{n \to \infty} [f^{(n)}(0)]^{1/n} = 0.$$

This condition on the coefficients of the expansion of an entire $f(z)$ in powers of $z$ is equivalent to condition (b) of §3 holding over the $z$-plane. Hence $f(z)$ is of type zero over the plane.

The scope of Theorem XX is clarified by the following two observations. First, $f(z)$ *may be entire, and of type zero in a sector of angle greater than* $\pi$, *and yet not of type zero over the plane*; secondly, $f(z)$ *may be a transcendental function of type zero over the plane without being of genus zero.* For the first case consider the function

$$f(z) = \int_0^\infty e^{-zt} e^{-t^2} dt.$$

Since the integral converges uniformly over every bounded region of the

$z$-plane, $f(z)$ is an entire analytic function of $z$. Now it can easily be seen that, if $t$ is considered as a complex variable, the positive half of the real $t$ axis, used in the above integral for contour, can be replaced by any half-line from the origin which makes an angle $\theta$ between $-\pi/4$ and $\pi/4$ with that positive $t$ axis. For any one $\theta$, we find

$$|f(z)| < \frac{A_\theta}{R(e^{i\theta}z) + B_\theta},$$

for $R(e^{i\theta}z) > -B_\theta$. By combining the inequalities for $\theta=\theta_1$ and $\theta=-\theta_1$ we thus easily see that $f(z)$ is in fact of order zero over any sector of angle less than $3\pi/2$ bisected by the positive half of the real $z$ axis. It is therefore also of type zero. That it is not of type zero over the plane is seen by considering negative real values of $z$, for which we have, with arbitrarily large $N$,

$$f(z) > \int_0^N e^{-zt}e^{-t^2}dt > \frac{e^{-N^2}}{(-z)}\,[e^{N(-z)} - 1].$$

For the second observation consider

$$f(z) = \prod_{n=2}^{\infty}\left[1 - \frac{z^2}{(n\log n)^2}\right],$$

whose zeros are $\pm n\log n$, $n=2, 3, \cdots$. As $\sum 1/(n\log n)$ does not converge, $f(z)$ is not of genus zero. On the other hand

$$|f(z)| < \prod_{n=2}^{\infty}\left[1 + \frac{|z|^2}{(n\log n)^2}\right] < P_N(|z|)\,\frac{\sinh(\pi|z|/\log N)}{\pi|z|/\log N},$$

where $P_N(|z|)$ is the polynomial in $|z|$ formed from the first $N-2$ factors. By expressing the hyperbolic sine in terms of exponentials, we see that condition (b) of §3 is satisfied over the plane, that is, $f(z)$ is of type zero over the plane.*

15. **The Laplace integral equation.** Except for a few results that followed directly from our definition of generalized differentiation, the preceding sections studied that definition for operators which we called of type zero. In the present and following sections, we shall make an independent study of operators given as Laplace integrals. We have seen in §11 that $A[f](t)$ satisfies the Laplace integral equation (48) when $f(D)$ is of order zero. We turn now to an extension of the converse of this result.

It has been shown in the literature that if $\int_a^\infty \psi(t)e^{-zt}dt$, where $\psi(t)$ is

* For the standard derivation of the same inequalities in the case of functions of genus zero see Borel, *Fonctions Entières*, chapter 3.

continuous for $t \geqq a$, converges for $\zeta = \zeta_0$, then it converges for all $\zeta$'s with $R(\zeta) > R(\zeta_0)$, and represents an analytic function of $\zeta$ in that half-plane.* This proof is readily extended to the case where $\psi(t)$ is continuous only for $t > a$, by breaking up the interval of integration. With this in mind we state

THEOREM XXI. *Let $\psi(t)$ be a continuous function of $t$ for $t > 0$, and let $\int_0^\infty \psi(t)e^{-\zeta t}dt$, considered improper at both its limits, converge for some value of $\zeta$. Then, if*

$$\int_0^\infty \psi(t)e^{-\zeta t}dt = f(\zeta)$$

*in the resulting half-plane of convergence, we will have, for $t > 0$,*

$$\psi(t) = A[f](t) = \lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \frac{(-1)^r f^{(r)}(1/\Delta x)}{r!\Delta x^{r+1}}. \tag{63}$$

Before we turn to the proof, note that when $f(D)$ is of order zero, $A[f](t)$ may be discontinuous for $t = 0$. Hence the assumption of continuity for $\psi(t)$ only for positive $t$. Such an assumption of mere continuity is in accord with the literature. Since we have observed that for $f(D)$ not only of order zero, but of type zero, $A[f](t)$ is analytic for positive $t$, we see that the present development must be a quite independent study of our fundamental definition.

The proof of the convergence theorem for Laplace integrals can easily be extended to show that successive derivatives of the integral can be found by differentiating under the integral sign. We therefore find, for the $f(\zeta)$ of our theorem,

$$A[f](r, \Delta x) = \frac{(-1)^r f^{(r)}(1/\Delta x)}{r!\Delta x^{r+1}} = \int_0^\infty \frac{e^{-t/\Delta x} t^r}{r!\Delta x^{r+1}} \psi(t)dt,$$

provided $1/\Delta x$ is in the half-plane of convergence. Let $t = r\Delta x \cdot \tau$. With the help of Stirling's formula for $r!$ we obtain

$$A[f](r, \Delta x) = \frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r} \int_0^\infty (e^{1-\tau}\tau)^r \psi(r\Delta x \cdot \tau)d\tau,$$

where $\epsilon_r \to 0$ as $r \to \infty$. The function $e^{1-\tau}\tau$ attains a maximum value of one for $\tau = 1$. Hence $(e^{1-\tau}\tau)^r$ stays equal to one for $\tau = 1$ as $r \to \infty$, but otherwise becomes inappreciable, as $r$ increases, except for an ever narrowing neighborhood of $\tau = 1$. This suggests that we break up the integral over limits $(0, 1-\lambda)$, $(1-\lambda, 1+\lambda)$, $(1+\lambda, \infty)$, where $\lambda$ is between zero and one, and

* For references, see Pincherle, loc. cit., p. 40.

show that the first and third parts can be neglected in finding the limit of $A[f](r, \Delta x)$.

With this in mind, let $b$ be a real number in the half-plane of convergence of the Laplace integral. That means, with our change of variable, that

$$\int_0^\infty e^{-b\cdot r\Delta x\cdot\tau}\psi(r\Delta x\cdot\tau)d\tau$$

converges. Now rewrite the integrand $(e^{1-\tau}\tau)^r\psi(r\Delta x\cdot\tau)$ so that it reads $(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r e^{-b\cdot r\Delta x\cdot\tau}\psi(r\Delta x\cdot\tau)$. For fixed $\lambda$, and $\Delta x$ sufficiently small, $(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r$ will monotonically increase from $\tau=0$ to $\tau=1-\lambda$, and monotonically decrease from $\tau=1+\lambda$ to $\tau=\infty$. We can therefore apply the second law of the mean for integrals,* and obtain, with $0\leqq\theta\leqq 1-\lambda$,

$$\int_0^{1-\lambda}(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r e^{-b\cdot r\Delta x\cdot\tau}\psi(r\Delta x\cdot\tau)d\tau$$
$$= [(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r]_{\tau\to+0}\int_0^\theta e^{-b\cdot r\Delta x\cdot\tau}\psi(r\Delta x\cdot\tau)d\tau$$
$$+ [(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r]_{\tau=1-\lambda}\int_\theta^{1-\lambda} e^{-b\cdot r\Delta x\cdot\tau}\psi(r\Delta x\cdot\tau)d\tau.$$

The first term of the right hand member is always zero. As for the second term, the integral, expressed in the original variable $t$, is

$$\frac{1}{r\Delta x}\int_{r\Delta x\cdot\theta}^{r\Delta x(1-\lambda)} e^{-bt}\psi(t)dt,$$

and, due to the convergence of $\int_0^\infty e^{-bt}\psi(t)dt$, stays bounded irrespective of its limits of integration, as $r\Delta x$ approaches a finite limit. On the other hand, as $\Delta x\to+0$, $e^{1-\tau+b\Delta x\cdot\tau}\tau$, for $\tau=1-\lambda$, becomes and remains less than a fixed positive quantity which is itself less than one. Since at the same time $r$ must increase indefinitely, we thus obtain

$$\lim_{\substack{\Delta x\to+0\\ r\Delta x\to t}}\frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r}[(e^{1-\tau+b\Delta x\cdot\tau}\tau)^r]_{\tau=1-\lambda}=0.$$

Hence the $(0, 1-\lambda)$ contribution to $A[f](r, \Delta x)$ can be neglected in studying its limit. An entirely similar proof shows the same to be true of the $(1+\lambda, \infty)$ contribution. We thus have, provided either limit exists,

* In the present instance we need the second law of the mean for an improper integral that may be only conditionally convergent. If, however, we first apply this law to the corresponding integral with positive lower limit $\epsilon$ (see de la Vallée Poussin, *Cours d'Analyse*, vol. 2, 1925, pp. 1–3) and if we then let $\theta$ be a limit value of the corresponding $\theta_\epsilon$'s as $\epsilon\to+0$, we easily obtain the desired result.

$$\lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} A[f](r, \Delta x) = \lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau}\tau)^r \psi(r\Delta x \cdot \tau) d\tau.$$

Since $(e^{1-\tau}\tau)^r$ is positive, and $\psi(r\Delta x \cdot \tau)$ is continuous in $\tau$, we have

$$\frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau}\tau)^r \psi(r\Delta x \cdot \tau) d\tau = \psi[r\Delta x(1+\theta\lambda)] \cdot \frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau}\tau)^r d\tau,$$

with $-1<\theta<1$. Now consider the special case $\psi(t)\equiv 1$. From the Laplace integral we find

$$f(\zeta) = \int_0^\infty e^{-\zeta t} dt = 1/\zeta.$$

For this $f$ direct calculation gives

$$A[f](r, \Delta x) = \frac{(-1)^r f^{(r)}(1/\Delta x)}{r!\Delta x^{r+1}} \equiv 1,$$

so that, from the above expression for $\lim A[f](r, \Delta x)$, we obtain

$$\lim_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} \frac{[r/(2\pi)]^{1/2}}{1+\epsilon_r} \int_{1-\lambda}^{1+\lambda} (e^{1-\tau}\tau)^r d\tau \equiv 1.$$

Combining these results, we thus see that

$$\limsup_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} A[f](r, \Delta x) \leqq M_\lambda, \quad \liminf_{\substack{\Delta x \to +0 \\ r\Delta x \to t}} A[f](r, \Delta x) \geqq m_\lambda,$$

where $M_\lambda$ and $m_\lambda$ are the upper and lower bounds of $\psi[t(1+\theta\lambda)]$ as $\theta$ varies from $-1$ to 1. Now let $\lambda \to +0$. The continuity of $\psi$ makes $M_\lambda$ and $m_\lambda$ both approach $\psi(t)$ as limit. Hence the upper and lower limits of $A[f](r, \Delta x)$ both equal $\psi(t)$, i.e., $A[f](t)$ exists, and equals $\psi(t)$.

16. **Operator expressed as a Laplace integral.** With the help of the preceding section we shall now prove

THEOREM XXII. *If $f(\zeta)$ is given by the convergent Laplace integral*

$$f(\zeta) = \int_0^\infty \psi(t) e^{-\zeta t} dt,$$

*where $\psi(t)$ is continuous for positive $t$, and the integral is considered improper at both its limits, and if $\phi(x)$ is continuous in the closed interval $(x_0, X)$, then*

$$\{f(D)\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X}\psi(X-x)\phi(x)dx, \tag{64}$$

*under either of the following two auxiliary conditions:*

(a) *the improper integral* $\int_0^h \psi(t)dt$ *is absolutely convergent for positive* $h$;

(b) $\phi(x)$ *has a finite total variation in a left neighborhood of* $x=X$.

It may be worth noting that (a) is the necessary and sufficient condition on a $\psi(t)$ continuous for positive $t$ that $\int_{x_0}^{X}\psi(X-x)\phi(x)dx$ converge for every continuous $\phi(x)$; while (b) is the necessary and sufficient condition on a continuous $\phi(x)$ that this integral converge for every $\psi(t)$ which is continuous for positive $t$, and for which $\int_0^h \psi(t)dt$ converges.

Since the preceding section proves that $A[f](t)$ exists, and equals $\psi(t)$, it is sufficient for the proof of our theorem, as in the treatment of operators of order zero given in §5, to show that

$$\lim_{h\to+0}\ \limsup_{\Delta x\to+0}\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x)\right| = 0.$$

We have by definition, and from the preceding section,

$$\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x) = \sum_{r=0}^{q}A[f](r,\Delta x)\phi(X-r\Delta x)\Delta x,\quad q = \{h/\Delta x\},$$

$$A[f](r,\Delta x) = \int_0^\infty \frac{e^{-t/\Delta x}t^r}{r!\Delta x^{r+1}}\psi(t)dt.$$

We shall now break up this integral over limits $(0, k)$, and $(k, \infty)$, where $k$ is greater than $h$ but independent of $r$ and $\Delta x$, and consider the two sums into which the above sum is thus broken up.

The treatment of the sum arising from the integrals with limits $(k, \infty)$ is independent of the special conditions (a), (b). Let

$$\alpha_r(t) = \frac{e^{-t/\Delta x}t^r}{r!\Delta x^{r+1}}.$$

We observe that $\alpha_r(t)$ is a monotonically decreasing function of $t$ in the interval $(k, \infty)$, due to the inequalities $k>h>r\Delta x$. The same will then be true of $\alpha_r(t)e^{bt}$ for sufficiently small $\Delta x$. As in the last section, the second law of the mean for integrals is applicable, and gives

$$\left|\int_k^\infty \alpha_r(t)\psi(t)dt\right| = \alpha_r(k)e^{bk}\left|\int_k^\mu e^{-bt}\psi(t)dt\right| \leqq N\alpha_r(k)e^{bk}.$$

The sum in question is thus not greater in absolute value than

$$MN\sum_{r=0}^{q}\alpha_r(k)e^{bk}\Delta x,$$

where $M$ is the upper bound of $|\phi(x)|$ in $(x_0, X)$. The above inequalities $k>h>r\Delta x$ also show that $\alpha_r(k)$ increases with $r$, as $r$ varies from 0 to $q$. Since $(q+1)\Delta x<h+\Delta x$, we find this sum to be less in absolute value than

$$(h+\Delta x)MN\alpha_q(k)e^{bk}.$$

As $\Delta x\to+0$, $q$ increases indefinitely, and so $\alpha_q(k)\to 0$. This sum can therefore be neglected in our discussion.

For condition (a) it will be sufficient to observe that the sum for the $(0, k)$ integrals will not exceed in absolute value

$$M\int_0^k\left(\sum_{r=0}^{q}\frac{e^{-t/\Delta x}t^r}{r!\Delta x^r}\right)|\psi(t)|\,dt.$$

But we have directly

$$\sum_{r=0}^{q}\frac{e^{-t/\Delta x}t^r}{r!\Delta x^r}<e^{-t/\Delta x}\sum_{r=0}^{\infty}\frac{t^r}{r!\Delta x^r}\equiv 1.$$

Hence this sum is less in absolute value than $M\int_0^k|\psi(t)|dt$, and hence also

$$\limsup_{\Delta x\to+0}\left|\left\{f\left(\frac{\Delta}{\Delta x}\right)\right\}_{X-h}^{X}\phi(x)\right|\leqq M\int_0^k|\psi(t)|\,dt.$$

Now let $h\to+0$, and at the same time let $k\to+0$, while keeping $k>h$. Due to the convergence of the last integral under condition (a), it will approach zero as limit with $k$, thus proving our result.

For condition (b), we shall write this sum in the form

$$\sum_{r=0}^{q}\phi(X-r\Delta x)\beta_r\Delta x,\quad \beta_r=\int_0^k\alpha_r(t)\psi(t)dt,$$

with $\alpha_r(t)$ as already defined, and use the identity

$$\sum_{r=0}^{q}\phi(X-r\Delta x)\beta_r\Delta x=\phi(X)\sum_{r=0}^{q}\beta_r\Delta x+[\phi(X-\Delta x)-\phi(X)]\sum_{r=1}^{q}\beta_r\Delta x$$
$$+\cdots+[\phi(X-q\Delta x)-\phi(X-[q-1]\Delta x)]\beta_q\Delta x.$$

We can write

$$\sum_{r=s}^{q}\beta_r\Delta x=\int_0^k\gamma_s(t)\psi(t)dt,\quad \gamma_s(t)=\sum_{r=s}^{q}\alpha_r(t)\Delta x.$$

Integration by parts gives the relation

$$\int_0^k \gamma_s(t)\psi(t)dt = \gamma_s(k)\int_0^k \psi(t)dt - \int_0^k \left[\int_0^t \psi(\tau)d\tau\right]\gamma_s'(t)dt.$$

Let $N_k$ be the upper bound of $\left|\int_0^t \psi(\tau)d\tau\right|$ in the interval $0 \leqq t \leqq k$. Note that $N_k$ is finite, since the convergence of the Laplace integral entails the convergence of this integral. We then have

$$\left|\int_0^k \gamma_s(t)\psi(t)dt\right| \leqq N_k\left[\gamma_s(k) + \int_0^k |\gamma_s'(t)|\,dt\right].$$

Now we easily verify that

$$\gamma_s(k) < 1, \quad \gamma_s'(t) = \alpha_{s-1}(t) - \alpha_q(t).$$

Furthermore, we have, for $\alpha_{s-1}(t)$, and similarly for $\alpha_q(t)$,

$$\int_0^k \alpha_{s-1}(t)dt < \int_0^\infty \alpha_{s-1}(t)dt \equiv 1,$$

so that we find

$$\left|\sum_{r=s}^{q}\beta_r\Delta x\right| = \left|\int_0^k \gamma_s(t)\psi(t)dt\right| < 3N_k.$$

Returning to our identity we thus obtain

$$\left|\sum_{r=0}^{q}\beta_r\phi(X - r\Delta x)\Delta x\right| < 3N_k[\,|\phi(X)| + V_{X-h}^{X}\phi(x)\,]$$

where $V_{X-h}^{X}\phi(x)$ is the total variation of $\phi(x)$ in $(X-h, X)$. Under hypothesis (b), this is finite for sufficiently small $h$, and so does not then increase as $h \to 0$. On the other hand, $N_k$ approaches zero as limit, as $k$, along with $h$, approaches zero as limit. Our theorem is thus proved under (b).

To illustrate (64), consider the operator $B(D, n)$. Through the change of variable $x = e^{-t}$, the usual integral for the beta function becomes

$$B(\zeta, n) = \int_0^\infty (1 - e^{-t})^{n-1}e^{-\zeta t}dt,$$

where $R(n) > 0$ (and $R(\zeta) > 0$). Since condition (a) is satisfied, we thus have, for $\phi(x)$ continuous in $(x_0, X)$,

$$\{B(D, n)\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X}[1 - e^{-(X-z)}]^{n-1}\phi(x)dx.^{*} \tag{65}$$

* (65) in conjunction with (44) leads to the solution of certain linear differential equations with exponential polynomial coefficients by means of definite integrals. Note, of course, that $B(D, n)$ is of order zero for $R(n) > 0$.

17. **Carson's form; $e^{-aD}$ as operator.** Carson has given a form for generalized differentiation which, with our notation, reads, *if*

$$\frac{f(\zeta)}{\zeta} = \int_0^\infty \chi(t)e^{-\zeta t}dt,$$

*then*

$$\{f(D)\}_{x_0}^{X}\phi(x) = \frac{d}{dX}\int_{x_0}^{X}\chi(X-x)\phi(x)dx.^* \tag{66}$$

We shall show that this form results from our definition in either of the following two cases:

(a) $\chi(t)$ *is continuous for* $t\geqq 0$, *with* $\lim_{t\to\infty}e^{-\zeta t}\chi(t)=0$ *for sufficiently large* $\zeta$; $\chi'(t)$ *exists for* $t>0$, *and, with* $\phi(x)$, *satisfies the hypothesis of Theorem* XXII.

(b) $\chi(t)$ *is continuous for* $t>0$, *with* $\int_0^k|\chi(t)|dt$ *convergent, and* $\phi'(x)$ *exists, and is continuous, for* $x_0\leqq x\leqq X$.

It may be noted that in case (a) Carson's form reduces to that of §16, while in case (b) it is related to that of §16 much as our operators of order one are related to those of order zero.

In case (a), let first $\chi(0)=0$. We have directly

$$\frac{d}{dX}\int_{x_0}^{X}\chi(X-x)\phi(x)dx = \int_{x_0}^{X}\chi'(X-x)\phi(x)dx.$$

On the other hand, integration by parts of the integral for $f(\zeta)/\zeta$ gives

$$f(\zeta) = \int_0^\infty \chi'(t)e^{-\zeta t}dt,$$

so that, by Theorem XXII, we have also

$$\{f(D)\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X}\chi'(X-x)\phi(x)dx.$$

(66) thus follows. If $\chi(0)=c$, we have

$$\int_0^\infty[\chi(t)-c]e^{-\zeta t}dt = \frac{f(\zeta)-c}{\zeta}.$$

The case just proved can therefore be applied to $f(D)-c$, so that

---

* J. R. Carson, *The Heaviside operational calculus*, Bulletin of the American Mathematical Society, vol. 32 (1926), pp. 43–68. Numerous papers on this subject by Carson are to be found in the Bell System Technical Journal.

$$\{f(D) - c\}_{x_0}^{X}\phi(x) = \frac{d}{dX}\int_{x_0}^{X}[\chi(X - x) - c]\phi(x)dx,$$

which easily reduces to (66)

In case (b), replace $X-x$ by a new variable. We thus get

$$\frac{d}{dX}\int_{x_0}^{X}\chi(X - x)\phi(x)dx = \int_{x_0}^{X}\chi(X - x)\phi'(x)dx + \chi(X - x_0)\phi(x_0)$$

where we returned to the old variable $x$ in the end. Now by §§15 and 16

$$A[u^{-1}f(u)](t) = \chi(t), \quad \{D^{-1}f(D)\}_{x_0}^{X}\phi'(x) = \int_{x_0}^{X}\chi(X - x)\phi'(x)dx.$$

If we turn then to the first existence proof for operators of finite order, with $\rho=1$, we see that it will hold here provided

$$\sum_{r=0}^{p} | A[u^{-1}f(u)](r, \Delta x) | \Delta x$$

is bounded as $\Delta x \to +0$. But the discussion of §16, with $u^{-1}f(u)$ in place of $f(u)$, shows that for $k > X - x_0$

$$\limsup_{\Delta x \to +0} \sum_{r=0}^{p} | A[u^{-1}f(u)](r, \Delta x) | \Delta x \leqq \int_{0}^{k} | \chi(t) | dt,$$

which is here assumed finite. Hence formula (17), with $\rho=1$, can be used to give

$$\{f(D)\}_{x_0}^{X}\phi(x) = \int_{x_0}^{X}\chi(X - x)\phi'(x)dx + \chi(X - x_0)\phi(x_0),$$

which is just what we found for the other member of (66).

More generally, if $\chi(t)$ and $\phi(x)$ satisfy the hypothesis of Theorem XXII, (66) is equivalent to the relation

$$\{f(D)\}_{x_0}^{X}\phi(x) = \frac{d}{dX}\{D^{-1}f(D)\}_{x_0}^{X}\phi(x),$$

which offers no difficulty when $f(D)$ is of finite order, or of type zero (see §7). However, the complete discussion of this relation when $f(\zeta)/\zeta$ is given by a Laplace integral offers considerable difficulty, and can only be made the subject of a separate investigation. We shall merely append an examination of the special operator $e^{-aD}$, $a>0$, which, in addition to its relation to the Laplace integral treatment, throws considerable light on our previous work with operators of type zero.

This operator comes under Carson's treatment by assigning to $\chi(t)$ the value 0, for $0 \leqq t \leqq a$, 1, for $t > a$. Since $\chi(t)$ is not even continuous for $t > 0$, agreement of the results with our definition cannot be sought for in the above two cases, but must be obtained directly. From our definition,

$$\left\{e^{-a(\Delta/\Delta x)}\right\}_{x_0}^{X}\phi(x) = \sum_{r=0}^{p} \frac{e^{-a/\Delta x}a^r}{r!\Delta x^r}\phi(X - r\Delta x), \quad p = \left\{\frac{X - x_0}{\Delta x}\right\}.$$

The coefficient of $\phi(X - r\Delta x)$ attains its largest value for $r\Delta x < a \leqq (r+1)\Delta x$. As in the treatment of the Laplace integral equation, it can be shown that only values of $r\Delta x$ in a neighborhood of this value affect the limit as $\Delta x \to +0$. We thus similarly obtain, for continuous $\phi(x)$,

(67) $X - x_0 < a$: $\left\{e^{-aD}\right\}_{x_0}^{X}\phi(x) = 0$; $X - x_0 > a$: $\left\{e^{-aD}\right\}_{x_0}^{X}\phi(x) = \phi(X - a)$,*

which agree exactly with Carson's results.† We may note in passing, that (31) gives the relation

$$A\left[e^{-au}f(u)\right](r, \Delta x) = \sum_{s=0}^{r} \frac{e^{-a/\Delta x}a^s}{s!\Delta x^s} A\left[f(u)\right](r - s, \Delta x),$$

from which, in a similar manner, we find

(68) $t < a$: $A\left[e^{-au}f(u)\right](t) = 0$; $t > a$: $A\left[e^{-au}f(u)\right](t) = A\left[f(u)\right](t - a)$,

provided $f(D)$ is of type zero. If we replace $f(u)$ by $u^{-1}f(u)$, we also obtain, by (12),

(69) $X - x_0 < a$: $\left\{e^{-aD}f(D)\right\}_{x_0}^{X}1 = 0$; $X - x_0 > a$: $\left\{e^{-aD}f(D)\right\}_{x_0}^{X}1 = \left\{f(D)\right\}_{x_0}^{X-a}1$.

Consider now some non-formal aspects of this and related operators. We found in §3 that, when $f(D)$ is of type zero, $A\left[f\right](t)$ exists for every positive $t$. By contrast, the formula

* For $X - x_0 = a$, terms on but one side of the maximum are included. Due to approximate symmetry of the coefficients with respect to this maximum, the result for $X - x_0 = a$ is seen to be $\phi(X - a)/2$.

† The reader may be interested in the following formal "derivation" of the formula of §16 from this formula. Writing symbolically

$$f(D) = \int_0^\infty \psi(t)e^{-tD}dt,$$

and noting that $\left\{e^{-tD}\right\}_{x_0}^{X}\phi(x)$ vanishes for $t > X - x_0$, we would be led to

$$\left\{f(D)\right\}_{x_0}^{X}\phi(x) = \int_0^\infty \psi(t)\left[\left\{e^{-tD}\right\}_{x_0}^{X}\phi(x)\right]dt = \int_0^{X-x_0} \psi(t)\phi(X - t)dt.$$

By replacing $X - t$ by $x$, we obtain the formula in question. A rigorous proof along these lines may be possible.

$$e^{-A}[e^{au}](r, \Delta x) = \frac{e^{-a/\Delta x}a^r}{r!\Delta x^r}$$

shows that $A[e^{-au}](t)$, while zero for every other positive $t$, fails to exist for $t=a$ through becoming infinite. Stronger still is the contrast furnished by $e^{aD}$, $a>0$, for which $A[e^{au}](t)$ fails to exist for every positive $t$ not exceeding a certain positive $\alpha$, and is zero for $t>\alpha$. Of course $e^{-aD}$ and $e^{aD}$ are not of type zero. They are however closely related to operators of type zero, since the corresponding analytic functions satisfy the analyticity condition (a), given in §3, for operators of type zero, and also the inequality of (b), not, however, for each positive $\kappa$, but only for $\kappa>\kappa_0$, $\kappa_0>0$. Calling operators satisfying (a), and this qualified (b), operators of type one, we easily find from §3 that, for them, $A[f](t)$ exists for $t$ greater than some positive $\alpha$. Also, by a modification of the treatment of part (b) of §6, we obtain the existence and expansion in series (21) of $\{f(D)\}_{x_1}^{X}\phi(x)$, provided $X-x_1$ is greater than this $\alpha$ and the radius of convergence of $\phi(x)$ at $x_1$ is greater than $X-x_1$ by more than a certain fixed positive $\beta$. These results were not included in the paper since the method used gave values to $\alpha$ and $\beta$ larger than those demanded by the operators themselves. They serve, however, to clarify the non-existence of $A[e^{-au}](t)$ and $A[e^{-au}](t)$ for certain $t$'s,* and the difference in form of $\{e^{-aD}\}_{x_0}^{X}\phi(x)$ for $X-x_0<a$, and $X-x_0>a$.

We may, in fact, think of our definition of generalized differentiation as not beginning to work, in the latter case, until $X-x_0>a$. As $a$ increases, this period of adjustment, as we may call it, increases. It is then interesting to observe that in the case of the operator $e^{-D^2}$ this period of adjustment is never completed, since, for finite $x_0$,

$$\{e^{-D^2}\}_{x_0}^{X}\phi(x) \equiv 0. \tag{70}$$

Here, then, we must take $x_0=-\infty$, for which, at least, (45) holds. $e^{-D^2}$ can be considered of type higher than one. This failure of our definition of generalized differentiation, with its partial failure in the case of operators of type one, throws into greater relief its peculiar applicability to those operators we have called of type zero.

---

* That $A[e^{-au}](t)$ and $A[e^{au}](t)$ are identically zero for $t>\alpha$ corresponds to $e^{-\alpha D}$ and $e^{\alpha D}$ being operators of type one over the plane.

NEW YORK CITY, N. Y.

# INTEGRAL SURFACES OF PAIRS OF PARTIAL DIFFERENTIAL EQUATIONS OF THE THIRD ORDER*

BY

ERNEST P. LANE

## I. INTRODUCTION

There is already in existence an extensive theory† of the integral surfaces of linear homogeneous partial differential equations of the *second order*. It is known that, by suitable choice of the parametric curves on an integral surface of a single equation of this kind, the equation can be reduced to an equation of Laplace when the surface sustains a conjugate net, or else to an equation of the parabolic type when the surface sustains a one-parameter family of asymptotic curves. Moreover, an integral surface of two essentially distinct equations of the second order is either a developable surface or else is a non-developable surface immersed in ordinary space $S_3$. Finally, an integral surface of three such equations is a plane.

It is natural to seek to develop a theory of integral surfaces of equations of the *third order* analogous to the theory briefly indicated in the foregoing paragraph. Some work in this direction‡ has already been done. On an integral surface $S$ of a single equation of the third order of the form

$$Ax_{uuu} + 3Bx_{uuv} + 3Cx_{uvv} + Dx_{vvv} + T = 0, \tag{1}$$

where not all of $A, B, C, D$ are zero and $T$ denotes, as throughout this paper, a linear combination of $x, x_u, x_v, x_{uu}, x_{uv}, x_{vv}$, there exists a covariantly defined triple§ of one-parameter families of curves whose differential equation is

$$Ddu^3 - 3Cdu^2dv + 3Bdudv^2 - Adv^3 = 0. \tag{2}$$

In fact, it will be shown in the next paragraph that at each point $P_x$ of the surface $S$ the triple of directions of these curves is the unique triple which is

---

* Presented to the Society, December 30, 1929; received by the editors November 20, 1929.

† Segre, *Su una classe di superficie degl'iperspazi legata colle equazioni lineari alle derivate parziali di 2° ordine*, Atti della Reale Accademia delle Scienze di Torino, vol. 42 (1907), p. 1047.

‡ Bompiani, *Determinazione delle superficie integrali d'un sistema di equazioni a derivate parziali lineari ed omogenee*, Rendiconti del Reale Istituto Lombardo di Scienze e Lettere, vol. 52 (1919), p. 626.

§ Segre, loc. cit., p. 1079.

apolar to every triple of directions of curves of section of $S$ and a suitably restricted variable hyperplane through $P_x$. There are different possibilities as to the coincidences of the families of curves (2), and as to the corresponding canonical forms that can be obtained for equation (1) by appropriate choices of the parametric curves on the surface $S$.

We shall now explain more precisely Segre's geometrical definition of the covariant triple of one-parameter families of curves represented analytically by equation (2). Let us consider a point $P_x$ on a surface $S$, and indicate as usual by the symbol $S$ (2, 0) the linear space of least dimensions that contains the osculating plane at the point $P_x$ of every curve on the surface $S$ that passes through $P_x$. It is well known that this space is determined by the six points $x, x_u, x_v, x_{uu}, x_{uv}, x_{vv}$, so that it is ordinarily a space of five dimensions. Any hyperplane with coördinates $\xi^{(1)}, \cdots, \xi^{(n+1)}$, which contains the space $S$ (2, 0), is restricted by the conditions

$$\text{(2a)}\quad \sum \xi x = 0,\ \sum \xi x_u = 0,\ \sum \xi x_v = 0,\ \sum \xi x_{uu} = 0,\ \sum \xi x_{uv} = 0,\ \sum \xi x_{vv} = 0,$$

the summation ranging from 1 to $n+1$. This hyperplane cuts the surface $S$ in a curve with a triple point at $P_x$, the directions of the triple-point tangents being given by the equation

$$\text{(2b)}\quad \sum \xi x_{uuu} du^3 + 3 \sum \xi x_{uuv} du^2 dv + 3 \sum \xi x_{uvv} du dv^2 + \sum \xi x_{vvv} dv^3 = 0.$$

But by means of the conditions (2a) we find from equation (1) the relation

$$A \sum \xi x_{uuu} + 3B \sum \xi x_{uuv} + 3C \sum \xi x_{uvv} + D \sum \xi x_{vvv} = 0,$$

which says that at the point $P_x$ the triple of directions given by equation (2) is the unique triple of directions which is *apolar* to every triple of directions given by equation (2b). Geometrically, if three directions (2b) coincide at all, they coincide in one of the three directions (2). Thus the origin of equation (2) and its geometrical significance are made clearer.

This paper is concerned with integral surfaces of a *pair* of linear homogeneous partial differential equations of the third order. The first problem is to reduce the pair of equations to canonical forms. It is found that there are six such canonical forms corresponding to six projectively distinct geometrically described classes of integral surfaces. The basis of the classification lies in the discussion of the singular triples of covariant families of curves in the next section, where the first canonical form is obtained. In the third section the other five canonical forms are reached. In the fourth section the integrability conditions are briefly discussed, and some geometric results are obtained.

## II. The first canonical form of the differential equations

In this section singular covariant triples of families of curves are defined on an integral surface of a pair of differential equations of the third order, and a beginning is made of the reduction of the pair of equations to canonical forms, the first canonical form being obtained.

Let us consider in a linear projective space of $n$ dimensions $S_n$ an integral surface $S$ of a pair of differential equations of the form

$$\begin{aligned} Ax_{uuu} + 3Bx_{uuv} + 3Cx_{uvv} + Dx_{vvv} + T = 0, \\ A'x_{uuu} + 3B'x_{uuv} + 3C'x_{uvv} + D'x_{vvv} + T' = 0. \end{aligned} \tag{3}$$

Let us suppose that these equations are essentially distinct in the sense that the respective coefficients of the third derivatives are not proportional. Let us further suppose that the surface $S$ is not an integral surface of more than two essentially distinct equations of the third order, and is not an integral surface of any equation of the second order.

The surface $S$ is an integral surface of a pencil* of differential equations. A general one of these can be written, on multiplying the first of equations (3) by a function $\lambda$ and the second by $\mu$, in the form

$$\alpha x_{uuu} + 3\beta x_{uuv} + 3\gamma x_{uvv} + \delta x_{vvv} + T = 0, \tag{4}$$

where

$$\alpha = \lambda A + \mu A', \quad \beta = \lambda B + \mu B', \quad \gamma = \lambda C + \mu C', \quad \delta = \lambda D + \mu D'. \tag{5}$$

On the surface $S$ there exists a system of covariant triples of one-parameter families of curves. This system depends linearly on the ratio $\lambda/\mu$, the differential equation of a general triple being, after the analogy of equaton (2),

$$\delta du^3 - 3\gamma du^2 dv + 3\beta du dv^2 - \alpha dv^3 = 0. \tag{6}$$

Such of these triples as consist of two coincident families and one distinct family will be called *singular triples of the first kind*; the two coincident families may be referred to as a *double family*, and the distinct family as a *single family*. Such of the triples as consist of three coincident families will be called *singular triples of the second kind*; the three coincident families may be referred to as a *three-fold family*. Those values of $\lambda/\mu$ which correspond to singular triples are roots of the quartic equation obtained by setting equal to zero the discriminant of the binary cubic form in equation (6), namely,

$$(\alpha\delta - \beta\gamma)^2 - 4(\alpha\gamma - \beta^2)(\beta\delta - \gamma^2) = 0. \tag{7}$$

* Bompiani, loc. cit., p. 630.

Furthermore, those values of $\lambda/\mu$ which correspond to singular triples of the second kind are roots of the three quadratic equations

$$\alpha\delta - \beta\gamma = 0, \quad \alpha\gamma - \beta^2 = 0, \quad \beta\delta - \gamma^2 = 0. \tag{8}$$

The differential equations (4) that correspond to values of $\lambda/\mu$ satisfying equation (7) may be called *singular equations.*

Since every algebraic equation has at least one root, it follows that *on every integral surface S of system* (3) *there is at least one singular triple of families of curves. Let us suppose in the remainder of this section that the surface S sustains no singular triple of the first kind. Then there exists on S at least one singular triple of the second kind.* Let us suppose that we have made a transformation of parameters so that the curves of this singular triple have become the $u$-curves. Now equation (6) shows that there exists a value of the ratio $\lambda/\mu$ such that $\beta=\gamma=\delta=0, \alpha\neq 0$. Let us take as the second of equations (3) the equation (4) that corresponds to this value of $\lambda/\mu$. Thus we see that the fundamental differential equations can be written in the form

$$\begin{aligned} 3Bx_{uuv} + 3Cx_{uvv} + Dx_{vvv} + T = 0, \\ x_{uuu} + T' = 0. \end{aligned} \tag{9}$$

Equations (5) now become

$$\alpha = \mu, \quad \beta = \lambda B, \quad \gamma = \lambda C, \quad \delta = \lambda D, \tag{10}$$

and therefore equation (7) reduces to

$$\lambda^2[B^2(4BD - 3C^2)\lambda^2 + 2C(2C^2 - 3BD)\lambda\mu + D^2\mu^2] = 0, \tag{11}$$

while equations (8) become

$$\lambda(\mu D - \lambda BC) = 0, \quad \lambda(\mu C - \lambda B^2) = 0, \quad \lambda^2(BD - C^2) = 0. \tag{12}$$

System (9) can be still further simplified. Every solution of equation (11) must, under the hypotheses that have been made, also satisfy equations (12). It is obvious that the solution $\lambda=0$ does so. Now it can be shown that $D\neq 0$ by showing that the supposition $D=0$ leads to a contradiction; the details of this indirect proof need not be elaborated here beyond saying that two cases may be considered according as $C=0$ or $C\neq 0$. Let us take $D=1$; the second factor of the left member of equation (11) vanishes only if $\lambda\neq 0$. The third of equations (12) gives $B=C^2$, and equation (11) now becomes

$$\lambda^2(C^3\lambda - \mu)^2 = 0,$$

equations (12) being all satisfied by the roots of this equation. Therefore *equations* (9) *can be written in the form*

$$3C^2x_{uuv} + 3Cx_{uvv} + x_{vvv} + T = 0, \qquad x_{uuu} + T' = 0. \tag{13}$$

*On an integral surface $S$ of this system the covariant triples have the differential equation*

$$\lambda(du - Cdv)^3 + (C^3\lambda - \mu)dv^3 = 0. \tag{14}$$

*Of these triples the only ones that are singular are the one for which $\lambda = 0$ and the one for which $C^3\lambda - \mu = 0$.*

In order to make the final step in the reduction of system (3) to the first canonical form, let us observe that the $v$-curves on the surface $S$ have not yet been specialized. Let us suppose that they are now chosen as the curves of the singular triple for which $C^3\lambda - \mu = 0$. This choice amounts to making $C = 0$ in equations (13). Our results may be summarized as follows.

*If the integral surfaces of system* (3) *have on them no singular triples of the first kind, then this system can be reduced to the canonical form*

$$x_{vvv} + T = 0, \qquad x_{uuu} + T' = 0, \tag{15}$$

*for which the differential equation of a general covariant triple is*

$$\lambda du^3 - \mu dv^3 = 0, \tag{16}$$

*and the singular triples are given by $\lambda^2\mu^2 = 0$.*

Incidentally it has been shown that *there are just two distinct singular triples of the second kind on an integral surface $S$ of a system of the form* (3) *which has on it no singular triple of the first kind.*

## III. Completion of the reduction to canonical forms

In this section the reduction of system (3) to canonical forms is continued and completed. Having already considered and disposed of the case in which an integral surface of system (3) has on it no singular triple of the first kind, we may now suppose that an integral surface $S$ has on it at least one singular triple of the first kind. Let us suppose that the parametric curves on the surface $S$ have been so chosen that the double family of a singular triple of the first kind consists of the $u$-curves and the single family consists of the $v$-curves. Now equation (6) shows that there exists a value of the ratio $\lambda/\mu$ such that $\alpha = \gamma = \delta = 0$, $\beta \neq 0$. Let us take the equation (4) that corresponds to this value of $\lambda/\mu$ as the second of equations (3). Thus we see that we can use as our fundamental equations the system

$$Ax_{uuu} + 3Cx_{uvv} + Dx_{vvv} + T = 0, \qquad 3x_{uuv} + T' = 0. \tag{17}$$

Equations (5) now become

$$\alpha = \lambda A, \quad \beta = \mu, \quad \gamma = \lambda C, \quad \delta = \lambda D, \tag{18}$$

and therefore equation (7) reduces to

$$\lambda[A(AD^2 + 4C^3)\lambda^3 - 6ACD\lambda^2\mu - 3C^2\lambda\mu^2 + 4D\mu^3] = 0, \tag{19}$$

while equations (8) become

$$\lambda(\lambda AD - \mu C) = 0, \quad \lambda^2 AC - \mu^2 = 0, \quad \lambda(\mu D - \lambda C^2) = 0. \tag{20}$$

It can be shown by an indirect proof that *there is another singular triple on the surface S besides the triple* $dudv^2=0$. In fact, if there were no other singular triple then $\lambda=0$ would be a quadruple root of equation (19). Consequently we should have $C=D=0$, $A\neq 0$. Taking $A=1$ we should find from equation (6) that every triple of the system

$$3\mu dudv^2 - \lambda dv^3 = 0$$

is singular, and a contradiction would thus be reached.

System (17) can be reduced still more by taking as the first equation therein another singular equation. Then $\mu=0$ is a solution of equation (19). It follows that

$$A(AD^2 + 4C^3) = 0. \tag{21}$$

Three possibilities present themselves:

1. $A = 0, \quad C \neq 0.$
2. $A = 0, \quad C = 0, \quad D \neq 0.$
3. $A \neq 0, \quad AD^2 + 4C^3 = 0.$

Let us for the moment suppose that the singular triple corresponding to $\mu=0$ is of the second kind. Since $\mu=0$ must be a solution of equations (20) it follows that $C=AD=0$, and two cases appear. First, if $C=A=0$, $D=1$, system (17) takes the canonical form

$$x_{vvv} + T = 0, \quad 3x_{uuv} + T' = 0, \tag{22}$$

for which the differential equation of the covariant triples is

$$\lambda du^3 + 3\mu dudv^2 = 0,$$

and the singular triples are given by $\lambda\mu^3=0$. Therefore *on an integral surface of system* (22) *there is one singular triple of the first kind and one of the second kind, so related that the curves of the three-fold family of the second triple are the*

*curves of the single family of the first triple.* In the second case, when $C=D=0$, $A=1$, the canonical form of system (17) is

$$x_{uuu}+T=0, \quad 3x_{uuv}+T'=0. \tag{23}$$

The differential equation of the covariant triples is

$$3\mu du dv^2-\lambda dv^3=0,$$

and *the singular triples are indeterminate,* every covariant triple being singular. *There are a one-parameter family of singular triples of the first kind and just one singular triple of the second kind on an integral surface of system* (23).

Let us finally suppose that the singular triple corresponding to $\mu=0$ is of the first kind. The possibilities not yet considered are the following:

1. $A=0, \quad C\neq 0.$

2. $ACD\neq 0, \quad AD^2+4C^3=0.$

It is necessary to consider two subcases in connection with the first possibility, according as $D=0$ or $D\neq 0$. In the first subcase, when $A=D=0$, $C=1$, the canonical form of system (17) is

$$3x_{uvv}+T=0, \quad 3x_{uuv}+T'=0. \tag{24}$$

The differential equation of the covariant triples is

$$\lambda du^2dv-\mu du dv^2=0,$$

and the singular triples are given by $\lambda^2\mu^2=0$. *On an integral surface of system* (24) *there are just two singular triples; both are of the first kind; the single family of each triple is the double family of the other.* In the second subcase, when $A=0$, $C=1$, $D\neq 0$, the canonical form of system (17) is

$$3x_{uvv}+Dx_{vvv}+T=0, \quad 3x_{uuv}+T'=0. \tag{25}$$

The differential equation of the covariant triples is

$$\lambda D du^3-3\lambda du^2dv+3\mu du dv^2=0,$$

and the singular triples are given by $\lambda\mu^2(3\lambda-4D\mu)=0$, their differential equations being, respectively,

$$du dv^2=0, \quad du^2(Ddu-3dv)=0, \quad du(2Ddu-3dv)^2=0.$$

*On an integral surface of system* (25) *there are just three singular triples; all are of the first kind; two of these triples have the same single family, which is also the double family of the other triple.*

For the last possibility, when $AC \neq 0$, $D=1$, $A=-4C^3$, the canonical form of system (17) is

$$-4C^3x_{uuu}+3Cx_{uvv}+x_{vvv}+T=0, \quad 3x_{uuv}+T'=0. \tag{26}$$

The differential equation of the covariant triples is

$$\lambda du^3-3\lambda C du^2dv+3\mu du dv^2+4\lambda C^3dv^3=0,$$

and the singular triples are given by $\lambda\mu=0$ and

$$\mu/\lambda=[3\pm 5(-15)^{1/2}]C^2/8,$$

their differential equations being, respectively,

$$\begin{gathered} du dv^2=0, \quad (du+Cdv)(du-2Cdv)^2=0, \\ \{4du+[1\mp(-15)^{1/2}]Cdv\}^2\{2du+[-7\pm(-15)^{1/2}]Cdv\}=0. \end{gathered} \tag{27}$$

*On an integral suface of system* (26) *there are four singular triples; all are of the first kind; there is no family common to any two of the singular triples.*

Thus the reduction of system (3) to canonical forms is completed. There are six of these forms written in equations (15), (22), (23), (24), (25), (26). Each canonical form has been characterized by a different projective geometric description of its integral surfaces, as may be clearly seen in the following way. Let us associate with each canonical form a symbol $(p, q)$ to indicate that on an integral surface there are $p$ singular triples of the first kind and $q$ of the second kind. Then the six canonical forms in the order designated above have, respectively, the symbols

$$(0,2),\ (1,1),\ (\infty,1),\ (2,0),\ (3,0),\ (4,0),$$

which are obviously all distinct.

It seems appropriate to say here precisely what is the relation between the present paper and the paper of Bompiani cited earlier. On page 630 of his paper Bompiani attacks the problem of reducing a pair of third-order equations to canonical forms, using apparently the same point of view that we have adopted. On page 631, he erroneously states* that "since the discriminant itself is of the third degree in the coefficients, therefore in the parameter, there will exist in general three equations in the pencil having the desired property." The words *third* and *three* should obviously be *fourth* and *four*, respectively. He obtains our equations (17), and then, appealing to some general theory developed earlier in his paper, arrives at two canonical forms which do not differ essentially from our canonical forms (23) and (24). He does not discover our four other canonical forms. He does not develop a

* Author's translation.

theory based on our canonical form (23), but confines his further studies to our canonical form (24), obtaining extremely interesting results to which we shall refer again in the next section.

## IV. Deductions from the canonical forms

For each of the six canonical forms of the differential equations under consideration there are certain integrability conditions that must be satisfied by the coefficients. The complete calculation of all of these would extend this paper beyond the bounds assigned for it. So the integrability conditions will only be discussed briefly in the first four cases here. The reader is referred for further details of these cases and the theory of the last two cases to the work of one of my students, Mr. C. F. Bowles, who is following up this study in his Chicago doctoral dissertation. Moreover, there are certain geometrical consequences easily deducible from the differential equations and the integrability conditions. Some of these geometrical results will also claim our attention in this section.

It is convenient to rewrite equations (15) in order to give explicit expressions for the terms of the second and lower orders:

$$
\begin{aligned}
x_{uuu} &= ax_{uu} + hx_{uv} + bx_{vv} + lx_u + mx_v + dx, \\
x_{vvv} &= a'x_{uu} + h'x_{uv} + b'x_{vv} + l'x_u + m'x_v + d'x.
\end{aligned} \tag{28}
$$

The form of these equations readily yields a geometric theorem, the statement of which is facilitated by Bompiani's *definition of the osculating space* $S(k, 0)$ *at a point* $P_x$ *of a surface. This space* $S(k, 0)$ *is the linear space of least dimensions containing the osculating linear space of* $k$ *dimensions* $S_k$ *at the point* $P_x$ *of every curve on the surface through* $P_x$. It follows from this definition that the space $S(k, 0)$ is determined by $x$ and the derivatives of $x$ up to and including those of order $k$, so that the space $S(k,0)$ is ordinarily of $\frac{1}{2}k \cdot (k+3)$ dimensions. In particular, as was pointed out in Section I, the space $S(2,0)$ at a general point of one of our surfaces is the space $S_5$ of the points

$$
x,\ x_u,\ x_v,\ x_{uu},\ x_{uv},\ x_{vv}. \tag{29}
$$

Now the form of the first of equations (28) shows* that *the osculating space* $S_3$ *at any point of a u-curve on an integral surface of system* (28) *lies in the space* $S(2, 0)$ *of the surface at the point.* The second of (28) gives a similar theorem regarding any $v$-curve on the surface.

We may be permitted to indicate here a couple of well known facts. The space $S(3, 0)$ at a point of a surface is ordinarily a space $S_9$. But at a point

* Bompiani, loc. cit., p. 628.

of an integral surface of two equations of the third order, such as we have under consideration in this paper, the space $S(3, 0)$ is a space $S_7$, since the two equations are two linear relations connecting the ten points determining the space $S(3, 0)$.

Let us consider the integrability conditions of equations (28). By means of the processes of differentiation and substitution it is possible to express each of the derivatives $x_{uuuu}, x_{uuuv}, x_{uvvv}, x_{vvvv}$ uniquely as a linear combination of $x_{uuv}, x_{uvv}$ and derivatives of lower orders including $x$ undifferentiated. Then it is possible to express each of the fifth derivatives of $x$ as a linear combination of the same derivatives and $x_{uuvv}$. Consequently the derivative $x_{uuuvvv}$ can be calculated in two ways as a linear combination of $x_{uuvv}$, $x_{uuv}$, $x_{uvv}$ and derivatives of lower orders. When these two expressions are equated, two cases arise, according as $a_v \neq b_u'$ or $a_v = b_u'$. In the first case, when $a_v \neq b_u'$, we obtain a differential equation of the fourth order in which the only fourth derivative is $x_{uuvv}$. Therefore, *if* $a_v \neq b_u'$, *every fourth derivative of* $x$ *can be expressed as a linear combination of* $x_{uuv}$, $x_{uvv}$ *and derivatives of lower orders, and an integral surface of system* (28) *is immersed in a space* $S_7$. The integrability conditions are to be found from the fact that each of the derivatives $x_{uuuvv}$ and $x_{uuvvv}$ can be calculated in two ways. In the second case, when $a_v = b_u'$, a differential equation of the third order results when the two expressions for $x_{uuuvvv}$ are equated. In virtue of assumptions already made, this equation must be an identity so that eight additional integrability conditions are found. Therefore, *if* $a_v = b_u'$, *every fifth derivative of* $x$ *can be expressed as a linear combination of* $x_{uuvv}$, $x_{uuv}$, $x_{uvv}$ *and derivatives of lower orders, and an integral surface of system* (28) *is immersed in a space* $S_8$.

Let us rewrite equations (22) in the form

$$
\begin{aligned}
x_{uuv} &= ax_{uu} + hx_{uv} + bx_{vv} + lx_u + mx_v + dx, \\
x_{vvv} &= a'x_{uu} + h'x_{uv} + b'x_{uv} + l'x_u + m'x_v + d'x.
\end{aligned} \tag{30}
$$

The second of these equations tells us again that the osculating space $S_3$ at each point of a $v$-curve is the space $S(2, 0)$ at the point.

In order to calculate the integrability conditions of system (30) one would first of all express $x_{uuuv}, x_{uuvv}, x_{uvvv}, x_{vvvv}$ as linear combinations of $x_{uuu}$, $x_{uvv}$ and derivatives of lower orders. Then $x_{uuuuv}, x_{uuuvv}, x_{uuvvv}, x_{uvvvv}$, $x_{vvvvv}$ can be expressed as linear combinations of $x_{uuuu}, x_{uuu}, x_{uvv}$ and derivatives of lower orders; moreover, $x_{uuvvv}$ can be so expressed in two ways, the coefficient of $x_{uuuu}$ being zero in one expression and being $a'$ in the other. When these two expressions are equated, two cases arise, according as $a' \neq 0$, or $a' = 0$. In the first case, when $a' \neq 0$, we obtain a differential equation

of the fourth order in which the only fourth derivative is $x_{uuuu}$. Therefore, *if* $a' \neq 0$, *every fourth derivative of* $x$ *can be expressed as a linear combination of* $x_{uuu}$, $x_{uvv}$ *and derivatives of lower orders, and an integral surface of system* (30) *is immersed in a space* $S_7$. The integrability conditions are to be found from two expressions for $x_{uuuuv}$. In the second case, when $a' = 0$, a differential equation of the third order results when the two expressions for $x_{uuvvv}$ are equated, which must be an identity, so that eight other integrability conditions are found, but the space of immersion of an integral surface of system (30) is not restricted to be other than a space $S_n$, with $n \geqq 7$.

Rewriting equations (23) in the form

$$\begin{aligned} x_{uuu} &= ax_{uu} + hx_{uv} + bx_{vv} + lx_u + mx_v + dx, \\ x_{uuv} &= a'x_{uu} + h'x_{uv} + b'x_{vv} + l'x_u + m'x_v + d'x \end{aligned} \tag{31}$$

we observe that the osculating space $S_3$ at each point of a $u$-curve lies in the space $S(2, 0)$ at the point. The derivatives $x_{uuuu}$, $x_{uuvv}$ can be expressed uniquely as linear combinations of $x_{uvv}$, $x_{vvv}$ and derivatives of lower orders, whereas $x_{uuuv}$ can be so expressed in two ways. Equating the two expressions for $x_{uuuv}$, we obtain the following eight integrability conditions:

$$\begin{aligned} b = 0,\ h = b',\ a_v = a_u' + a'h' + l',\quad & h_v + ah' + l = h_u' + a'h + h'^2 + m', \\ ab' + m = b_u' + h'b',\quad & l_v + al' = l_u' + a'l + h'l' + d', \\ m_v + am' + d = m_u' + a'm + h'm',\quad & \\ d_v + ad' = d_u' + a'd + h'd'. & \end{aligned} \tag{32}$$

It is not possible to express, by the processes of differentiation and substitution, all derivatives of the fifth or any higher order in terms of derivatives of lower order. Therefore an integral surface of system (31) may lie in any space $S_n$ with $n \geqq 7$.

Equations (24) may be rewritten in the form

$$\begin{aligned} x_{uuv} &= ax_{uu} + hx_{uv} + bx_{vv} + lx_u + mx_v + dx, \\ x_{uvv} &= a'x_{uu} + h'x_{uv} + b'x_{vv} + l'x_u + m'x_v + d'x. \end{aligned} \tag{33}$$

This system of equations together with the integrability conditions and the integral surfaces thereof have been studied by Bompiani,* who has considered the analogy between this system of two equations of the third order and the Laplace equation of the second order, and has studied particularly transformations of this system analogous to the well known transformations of Laplace for the equation that bears his name. Equating the two expressions

* Bompiani, loc. cit., pp. 632–636.

for $x_{uuvv}$ obtainable for equations (33) leads to eight integrability conditions of which two are $a'=b=0$. The space in which the integral surfaces lie is any space $S_n$ with $n \geqq 7$.

There is an interesting interpretation of the conditions $a'=b=0$ which seems to have escaped notice hitherto. Let us consider the ruled surface $R_u$ of $u$-tangents constructed at points of a $v$-curve. Let us consider the linear space of least dimensions which contains the osculating planes, at points of a generator $xx_u$ of the surface $R_u$, of every curve on $R_u$ that intersects the generator $xx_u$. This space is determined by the points $x$, $x_v$, $x_{vv}$, $x_u$, $x_{uv}$, $x_{uvv}$ and is ordinarily a space $S_5$ called the tangent space $S_5$ of the ruled surface $R_u$ along the generator $xx_u$. But the last of equations (33) shows that for an integral surface of these equations this space is a space $S_4$. Therefore, *for an integral surface of equations* (33), *the osculating planes, at points on a generator $xx_u$ on a ruled surface $R_u$, of all the curves on $R_u$ that intersect the line $xx_u$ lie in a space $S_4$.* There is a symmetric theorem with $u$ and $v$ interchanged.

The analogy which interested Bompiani is further illuminated by the following consideration. One of my students, Mr. G. D. Gore, has remarked to me that if two surfaces have their points in one-to-one correspondence so that their tangent planes at corresponding points intersect in straight lines, then each surface is an integral surface of a system of equations of the type (33), or else of an equation of Laplace, provided that the parameters of the surfaces are suitably chosen.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

# ON THE OVERCONVERGENCE OF SEQUENCES OF POLYNOMIALS OF BEST APPROXIMATION*

BY

J. L. WALSH

1. **Introduction.** There are many cases where the sequence of polynomials of best approximation to an analytic function in a given region $C$ converges to that function (or its analytic extension) not merely in $C$ but in a larger region containing $C$ in its interior. It is the object of the present paper to show that many such results can be derived easily from a single general theorem on the degree of approximation to an analytic function. We shall discuss also the corresponding facts for approximation to harmonic functions by harmonic polynomials; the analogy between the two theories is close. There are a number of cases to be treated, according as *best approximation* is measured (1) in the sense of Tchebycheff, (2) by line integrals taken over the curve bounding the region considered, (3) by line integrals taken over the unit circle in a new plane, after conformal mapping of the original region onto the interior of this circle, or (4) by surface integrals over the original region.

The term *overconvergence* is used to denote the behavior mentioned above, that although a given function is defined and approximated in a certain region, the sequence of polynomials approximating the function in the given region may converge uniformly in a larger region containing the given region in its interior. The term overconvergence has recently been used by Ostrowski in a somewhat different connection. Our primary topic is the overconvergence of sequences of polynomials, but we devote some space (§§11, 12) also to approximation where overconvergence need not take place.

THEOREM I. *Let $C$ be an arbitrary closed limited point set (not a single point) of the $z$-plane, whose complementary set with respect to the entire plane is simply connected. Let $w=\phi(z)$ be a function which maps conformally the exterior of $C$ onto the exterior of the unit circle in the $w$-plane so that the points at infinity correspond to each other. Let $C_R$ denote the curve $|\phi(z)|=R>1$, that is, the transform in the $z$-plane of the circle $|w|=R$.*

* Presented to the Society, February 22, 1930; received by the editors April 7, 1930. An abstract of this paper was published in the Proceedings of the National Academy of Sciences, vol. 16 (1930), p. 297.

*If the function $f(z)$ is analytic on and within $C_R$, then there exist polynomials $p_n(z)$ of respective degrees $n$, $n=0, 1, 2, \cdots$, such that the inequalities*

$$|f(z) - p_n(z)| \leqq \frac{M}{R^n}, \quad R > 1, \tag{1}$$

*where $M$ is independent of $n$ and of $z$, are valid for every $z$ in $C$.*

*If the polynomials $p_n(z)$ are given so that (1) is valid for every $z$ in $C$, then the sequence $\{p_n(z)\}$ converges everywhere interior to $C_R$ and uniformly on any closed point set interior to $C_R$, and thus $f(z)$ is analytic throughout the interior of $C_R$.**

**2. Tchebycheff polynomials.** The Tchebycheff polynomial of degree $n$ for approximation to $f(z)$ on $C$ is the polynomial $\pi_n(z)$ of degree $n$ such that

$$\max |f(z) - \pi_n(z)|, \quad z \text{ on } C,$$

is less than

$$\max |f(z) - p_n(z)|, \quad z \text{ on } C,$$

for any other polynomial $p_n(z)$ of degree $n$. The Tchebycheff polynomial exists and is unique.† The present writer has already proved elsewhere‡ the following theorem, but the proof is so short and so typical of the reasoning used later that it will be given here. The theorem was first proved by Faber§, using an entirely different method, and only in the restricted case that $C$ is bounded by an analytic Jordan curve.

THEOREM II. *Let $C$ and $C_R$ have the same significance as in Theorem* I. *If the function $f(z)$ is analytic interior to $C_R$, then the sequence $\{\pi_n(z)\}$ of Tchebycheff polynomials for approximation to $f(z)$ on $C$ converges to $f(z)$ interior to $C_R$, uniformly on an arbitrary closed point set interior to $C_R$.*

By the first part of Theorem I, if $R_1$ be an arbitrary number less than $R$, there exist polynomials $p_n(z)$ of respective degrees $n$ such that

$$|f(z) - p_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } C.$$

---

* Various parts of Theorem I are due to various writers, notably S. Bernstein, M. Riesz, Faber, Fjeér, Szegö. See Walsh, Münchner Berichte, 1926, pp. 223–229. An arbitrary function of the form $a_0+a_1z+\cdots+a_nz^n$ is said to be a polynomial of degree $n$.

† Tonelli, Annali di Matematica, (3), vol. 15 (1908), pp. 47–119.

‡ See Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 499–544, especially pp. 513, 514, and also *Boundary values of an analytic function and the Tchebycheff method of approximation*, these Transactions vol. 32 (1930), pp. 335–390.

§ Crelle's Journal, vol. 150 (1920), pp. 79–106; especially p. 105.

This inequality, holding for the polynomials $p_n(z)$, must also hold for the Tchebycheff polynomials:

$$|f(z) - \pi_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } C,$$

and since $R_1 < R$ is arbitrary, Theorem II follows from the second part of Theorem I.

If $C'$ is an arbitrary closed limited region, we denote by $C$ the totality of closed limited regions into which the boundary of $C'$ divides the plane; the boundary points of $C$ (considered as a set of regions, not as a point set) are precisely the boundary points of $C'$, and if inequality (1) holds for all points of $C'$, or even for all points of the boundary of $C'$, then (1) holds also for all points of $C$. The curve $C_R$ is defined as in Theorem I from this point set $C$, by means of the mapping of the complement of $C$. Thus, *if the function $f(z)$ is analytic interior to $C_R$, then the sequence $\{\pi_n(z)\}$ of Tchebycheff polynomials for approximation to $f(z)$ on $C'$ converges to $f(z)$ interior to $C_R$, uniformly on an arbitrary closed point set interior to $C_R$.* It will be noticed, in fact, that the Tchebycheff polynomial is the same for approximation to $f(z)$ on $C$, or on $C'$, or on the boundary of $C$ or $C'$. The function in the left-hand member of (1) has one and the same maximum for all these point sets.

Throughout the present paper we shall use the notation $C_R$ in the following sense. Let $C$ be an arbitrary closed limited point set and $D$ the set of points which can be joined to the point at infinity by Jordan arcs which do not meet $C$. Then the image of $|w| = R > 1$ when $D$ is mapped onto the exterior of $|w| = 1$ in the $w$-plane so that the points at infinity correspond to each other shall be denoted by $C_R$.

3. **Tchebycheff approximation with a norm function.** The notion of approximation by the method of Tchebycheff with a norm function will now be introduced. Let the function $p(z)$ be given continuous and different from zero on $C$. Then a Tchebycheff polynomial for approximation to $f(z)$ on $C$ with the norm function $p(z)$ is a polynomial $\pi_n'(z)$ of degree $n$ such that

$$\max [|p(z)||f(z) - \pi_n'(z)|], \quad z \text{ on } C,$$

is no greater than

$$\max [|p(z)||f(z) - p_n'(z)|], \quad z \text{ on } C,$$

for any other polynomial $p_n'(z)$ of degree $n$. The existence of the polynomial

$\pi_n'(z)$ can easily be proved* and the uniqueness follows from the original method of Tonelli.† *If $C$ and $C_R$ have the same significance as in Theorem* I *and if $f(z)$ is analytic within $C_R$, then the sequence of Tchebycheff polynomials $\{\pi_n'(z)\}$ for approximation to $f(z)$ on $C$ with an arbitrary norm function $p(z)$ converges to $f(z)$ interior to $C_R$, uniformly on an arbitrary closed point set interior to $C_R$.*

If $R_1$ is an arbitrary number less than $R$, the first part of Theorem I informs us that a sequence of polynomials $p_n(z)$ of respective degrees $n$ exists so that

$$|f(z) - p_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } C.$$

If we have $|p(z)| \leqq P$ for $z$ on $C$, we have also

$$|p(z)||f(z) - p_n(z)| \leqq \frac{MP}{R_1^n}, \quad z \text{ on } C,$$

from which follows

$$|p(z)||f(z) - \pi_n'(z)| \leqq \frac{MP}{R_1^n}, \quad z \text{ on } C.$$

If we have $|p(z)| \geqq q > 0$, $z$ on $C$, it follows that

$$|f(z) - \pi_n'(z)| \leqq \frac{MP}{qR_1^n}, \quad z \text{ on } C,$$

and our result is an immediate consequence of the second part of Theorem I.

If we consider the ordinary Tchebycheff polynomial $\pi_n(z)$, it is of course a matter of indifference whether we consider approximation to $f(z)$ on $C$, or approximation to $f(z)$ on the boundary $\Gamma$ of $C$, for the function in the left-

* The proof as given in the case of reals for the ordinary Tchebycheff polynomial applies without essential change. See for instance de la Vallée Poussin, *Approximation des Fonctions*, Paris, 1919, p. 74.

† Loc. cit. The principal modifications to be made in his treatment are the following (in his notation, p. 109, except for the norm function $n(x)$):

$$\mu = \max |Y(x)n(x)|, \; P_n(x_r) = Y(x_r).$$

Choose $\epsilon < \mu/2$ and $\delta$ so that $|x' - x''| < \delta$ implies

$$|Y(x')n(x') - Y(x'')n(x'')| < \epsilon, \; |P(x')n(x') - P(x'')n(x'')| < \epsilon.$$

Choose $\omega$ so that, in the circles $|x_r - x| \leqq \delta$, we have

$$|Y(x) - \omega P_n(x)| \; |n(x)| < \mu,$$

and so that on the entire point set $A$ we have

$$\omega |P_n(x)| \; |n(x)| < \frac{\mu - \mu''}{2},$$

where in $A'$

$$|Y(x)n(x)| \leqq \mu'' < \mu.$$

hand member of (1) has its maximum in $C$ on $\Gamma$. A similar remark holds for the polynomial $\pi_n'(z)$, if the norm function $p(z)$ is analytic on $C$.

4. **Approximation in the sense of least $p$th powers.** We turn now to consideration of polynomials of best approximation in the sense of least $p$th powers. Let $p$ be an arbitrary positive number greater than unity,* and let the point set $C$ now be the closed interior of a rectifiable Jordan curve $\Gamma$. Let $n(s)$ be an arbitrary function continuous and positive on $\Gamma$, where $s$ indicates arc length on $\Gamma$; this weight or norm function is usually chosen as unity, but that is not necessary. A polynomial of best approximation is a polynomial $\pi_n(z)$ of degree $n$ such that

$$\int_\Gamma |f(z) - \pi_n(z)|^p n(s)ds$$

is not greater than the corresponding expression

$$\int_\Gamma |f(z) - p_n(z)|^p n(s)ds$$

formed for any other polynomial $p_n(z)$ of degree $n$. The polynomial $\pi_n(z)$ can easily be shown to exist; its uniqueness follows from the method recently used by Julia† in a similar situation.

THEOREM III. *Let the function $f(z)$ be analytic within $C_R$, where the point set $C$ is the interior of the rectifiable Jordan curve $\Gamma$. Then the sequence of polynomials $\{\pi_n(z)\}$ of best approximation to $f(z)$ on $C$ in the sense of least weighted $p$th powers as measured on $\Gamma$ converges to $f(z)$ throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

We shall have occasion to use the following well known inequality for arbitrary continuous functions $F(x)$ and $G(x)$:

$$(2) \qquad \left| \int [F(x)]^\alpha [G(x)]^{1-\alpha} dx \right| \leqq \left( \int |F(x)|\, dx \right)^\alpha \left( \int |G(x)|\, dx \right)^{1-\alpha}.$$

If $R_1$ is an arbitrary number less than $R$, there exist by Theorem I polynomials $p_n(z)$ of respective degrees $n$ such that

$$|f(z) - p_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } \Gamma.$$

From this it follows that $M'$ exists such that

* Much of our work is valid if $p$ is assumed merely positive.

† Acta Litterarum ac Scientiarum (Szeged), vol. 14 (1929), pp. 217–226.

$$\int_\Gamma |f(z) - p_n(z)|^p n(s)ds \leqq \frac{M'}{R_1^{np}},$$

whence we deduce the corresponding inequality for the polynomials $\pi_n(z)$ of best approximation:

$$\int_\Gamma |f(z) - \pi_n(z)|^p n(s)ds \leqq \frac{M'}{R_1^{np}}. \tag{3}$$

The function $n(s)$, being positive and continuous on $\Gamma$, has a positive minimum $n'$ on $\Gamma$, so we may write

$$\int_\Gamma |f(z) - \pi_n(z)|^p ds \leqq \int_\Gamma |f(z) - \pi_n(z)|^p \frac{n(s)}{n'} ds \leqq \frac{M'}{n'R_1^{np}} = \frac{M_1'}{R_1^{np}}. \tag{4}$$

Here we apply Cauchy's integral formula

$$f(z) - \pi_n(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(t) - \pi_n(t)}{t - z} dt,$$

and use (4) in conjunction with (2), setting $\alpha = 1/p$, $F(x) = |f(t) - \pi_n(t)|^p$, $G(x) = 1/|t-z|^{p/(p-1)}$, $dx = |dt|$. If we so restrict $z$ as to lie on an arbitrary closed point set interior to $\Gamma$, we have then

$$|f(z) - \pi_n(z)| \leqq \frac{M'''}{R_1^n}, \tag{5}$$

where $M'''$ depends on the particular point set to which $z$ is restricted.

For definiteness we consider (5) to hold on an arbitrary Jordan curve $\Gamma'$ interior to $\Gamma$. The sequence $\{\pi_n(z)\}$ then converges, by the second part of Theorem I, interior to the curve $C_{R_1}'$, and uniformly on any closed point set interior to $C_{R_1}'$, where $C_{R_1}'$ is the image in the $z$-plane of the circle $|w| = R_1$ when the exterior of $\Gamma'$ is mapped onto the exterior of the unit circle in the $w$-plane so that the points at infinity correspond to each other. When $\Gamma'$ approaches $\Gamma$ uniformly, the curve $C_{R_1}'$ approaches $C_{R_1}$ uniformly.* Since $R_1$(less than $R$) is arbitrary, and since the convergence of the sequence $\{\pi_n(z)\}$ is uniform on any closed point set interior to $C_{R_1}'$, Theorem III is established.

Theorem III was proved for the case $p = 2$, $n(s) \equiv 1$ by Szegö† for the case that $\Gamma$ is an analytic curve and by Smirnoff‡ for the case that $\Gamma$ is the most

* Carathéodory, Mathematische Annalen, vol. 72 (1912), pp. 107–144, especially pp. 126, 127.

† Mathematische Zeitschrift, vol. 9 (1921), pp. 218–270.

‡ Journal de la Société Physico-Mathématique de Léningrade, vol. 2 (1928), pp. 155–178, especially pp. 177, 178.

general rectifiable curve. Theorem III was also proved by Szegö for the case $p=2$ and $\Gamma$ a circle.* In all these cases the method was different from the present one. Indeed, under the present hypothesis, the polynomial of best approximation to an arbitrary function $f(z)$ is of the form

$$\pi_n(z) = a_0P_0(z) + a_1P_1(z) + \cdots + a_nP_n(z),$$

where the polynomials $P_i(z)$ depend on $\Gamma$ but not on $f(z)$, and where the coefficients $a_i (i \leqq n)$ are independent of $n$. The method used by both Szegö and Smirnoff is the study of the asymptotic properties of the polynomials $P_i(z)$ and of the coefficients $a_i$. The first part of Theorem I is used by both writers.

A less specific result than Theorem III for certain more restricted types of curves has recently been established by Dunham Jackson.†

We have proved actually more than is stated in Theorem III, for we have used in the proof only inequality (3), by way of hypothesis on the polynomials $\pi_n(z)$.

5. **Least $p$th powers and conformal mapping; Jordan regions.** There is a result analogous to Theorem III which applies even if the curve $\Gamma$ is not rectifiable. This consists in measuring approximation not on $\Gamma$ but on $\gamma$, the unit circle in the $w$-plane, after mapping the interior of $\Gamma$ onto the interior of $\gamma$, setting $w=\phi(z)$, $z=\psi(w)$. This method has recently been used with success by Julia,‡ in the analogous situation of approximation of harmonic functions by harmonic polynomials. Indeed, the present paper would probably not have been written without the impetus due to his paper.

Let the function $n(\theta)$ be positive and continuous on $\gamma$, where $\theta$ is arc length on $\gamma$; we consider as the polynomial of degree $n$ of best approximation the polynomial $\pi_n(z)$ of degree $n$ such that

$$\int_\gamma | f(z) - \pi_n(z) |^p \, n(\theta)d\theta, \qquad z = \psi(w), \tag{6}$$

is less than the corresponding expression for any other polynomial of degree $n$, where $f(z)$ is given analytic on and within $\Gamma$. The notation $\gamma$ is to be used later as well; by measure of approximation with least $p$th powers on $\gamma$ we shall understand the integral (6) or (for harmonic functions) the integral analogous to it; the given region $C$ is to be mapped conformally on the

* Mathematische Zeitschrift, vol. 6 (1920), pp. 167–120.

† *On certain problems of approximation in the complex domain*, presented to this Society December 28, 1927, abstract published in the Bulletin of the American Mathematical Society, vol. 34 (1928,) p. 151; the paper is expected to appear in the same Bulletin. The abstract does not state the principal results of the paper, and that paper and the present one were written entirely independently of each other. The methods used are different in the two cases.

‡ Loc. cit. An arbitrary point interior to $\Gamma$ is made to correspond to the center of $\gamma$.

interior of $\gamma$. Theorem IVa will be proved by almost precisely the method used in the proof of Theorem III. As in Theorem III, the polynomial $\pi_n(z)$ exists and is unique; this may be proved by the method of Julia.

THEOREM IVa. *Let the function $f(z)$ be analytic within $C_R$, where $C$ is the interior of the Jordan curve $\Gamma$. Then the sequence of polynomials $\{\pi_n(z)\}$ of best approximation to $f(z)$ on $C$ in the sense of least weighted $p$th powers as measured on $\gamma$ converges to $f(z)$ throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

We can proceed as in the proof of Theorem III, now (as in expression (6) ) making frequent use of the theorem that the mapping of the interior of $\Gamma$ onto the interior of $\gamma$ is continuous in the closed regions. Inequality (3) holds, where the integral is now taken over $\gamma$ instead of $\Gamma$; the values of the integrand on $\gamma$ are the values of the functions at the corresponding points of $\Gamma$. Cauchy's integral formula is likewise valid, the integral being taken over $\gamma$, and hence (5) holds for $w$ on any closed point set interior to $\gamma$, that is, for $z$ on any closed point set interior to $\Gamma$; the number $M'''$ depends naturally on this point set. This point set can be chosen the closed interior of a Jordan curve $\Gamma'$ interior to $\Gamma$, so by the reasoning previously given based on the results of Carathéodory, Theorem IVa now follows.

The case of Theorem IVa for $p=2$, $n(\theta)\equiv 1$, is of particular interest, for in this case approximation is in the sense of least squares on $\gamma$, and the polynomial $\pi_n(z)$ is of the form*

$$\pi_n(z) = a_0P_0(z) + a_1P_1(z) + \cdots + a_nP_n(z),$$

where the polynomials $P_i(z)$ depend on the region $C$ and not on the function $f(z)$ and where the coefficients $a_i (i \leqq n)$ are independent of $n$. This yields, then, a sequence of polynomials $\{P_n(z)\}$ in terms of which can be developed an arbitrary function $f(z)$ analytic on and within $\Gamma$, and the development

$$a_0P_0(z) + a_1P_1(z) + \cdots + a_nP_n(z) + \cdots$$

converges uniformly to the function $f(z)$ on and within $\Gamma$. This problem *for the most general Jordan curve* $\Gamma$ was long ago proposed by Faber†; a solution was more recently announced by Fejér,‡ but so far as the present writer is aware the details have not yet been published.

These polynomials $\{P_n(z)\}$ may be obtained by starting with the special polynomials $\{z^n\}$, and orthogonalizing and normalizing this latter set with

* Compare Szegö, loc. cit.

† Mathematische Annalen, vol. 57 (1903), pp. 389–408.

‡ Göttinger Nachrichten, 1918, pp. 319–331.

respect to integration on $\gamma$. A corresponding remark holds naturally in every case below, for $p=2$, $n(s)\equiv 1$.

Theorem IVa is closely related to Theorem III, for there is a relation between arc length on $\Gamma$ (if $\Gamma$ is rectifiable) and arc length on $\gamma$, and the functions $n(s)$ and $n(\theta)$ are arbitrary. The writer knows, however, no proof that Theorem III is contained in Theorem IVa.

6. **Least $p$th powers and conformal mapping; general regions.** Theorem IVa is in reality but a special case, proved first because of its simplicity, of a more general theorem:

THEOREM IV. *Let $C$ be an arbitrary closed limited simply connected region and let the function $f(z)$ be analytic interior to $C_R$. Then the sequence of polynomials $\{\pi_n(z)\}$ of best approximation to $f(z)$ on $C$ in the sense of least weighted $p$th powers as measured on $\gamma$ converges to $f(z)$ throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

Under the conformal map $w=\phi(z)$, $z=\psi(w)$ of $C$ onto the interior of $\gamma$, almost all (i.e. all, except possibly for points of a set of measure zero) points of $\gamma$ correspond to accessible boundary points of $C$. For, as Carathéodory has pointed out, the function $\psi(w)$ is limited interior to $\gamma$, and hence by Fatou's theorem $\lim_{r\to 1, r<1}\psi(re^{i\theta})$ exists for almost all values of $\theta$. A radius $0\leqq r\leqq 1$, $\theta$ constant, for which this limit exists therefore corresponds under the conformal map to a Jordan arc from the point $\psi(0)$ to a boundary point of $C$. When this limit exists for a particular value of $\theta$, we define $\psi(e^{i\theta})$ as this limit.

The functions $f(z)$, $p_n(z)$ (about to be introduced), $\pi_n(z)$ are to be defined on $\gamma$ after conformal transformation by means of their limiting values by radial approach to $\gamma$; these limiting values exist for almost all values of $\theta$, by Fatou's theorem. Moreover, for any point of $\gamma$ corresponding to an accessible point $z'$ of the boundary of $C$, and thus for almost all points of $\gamma$, these limiting values exist and coincide with the values of the functions $f(z)$, $p_n(z)$, $\pi_n(z)$ at the point $z'$. For radial approach to $w=\phi(z')$ corresponds to approach to $z'$ along a Jordan arc, and for such approach the functions involved have limiting values, $f(z')$, $p_n(z')$, $\pi_n(z')$.

Cauchy's integral formula applies in the $w$-plane to any function such as $f(z)$, $p_n(z)$, $\pi_n(z)$, the integral now being taken in the sense of Lebesgue. We have, for instance,

$$f[\psi(w)] = \frac{1}{2\pi i}\int_{\gamma'} \frac{f[\psi(t)]}{t-w}dt, \quad t=\rho e^{i\theta},$$

where $\gamma'$ is the circle $|w|=\rho<1$. Keep $w$ fixed. When $\rho$ approaches unity

through a sequence of values, the integrand, considered as a function of $\theta$, approaches the limit $f[\psi(t)]/(t-w)$ for almost all values of $\theta$. During this approach of $\rho$ to the limit unity the integrand remains uniformly limited. It follows that Cauchy's formula

$$f[\psi(w)] = \frac{1}{2\pi i}\int_\gamma \frac{f[\psi(t)]}{t-w}dt$$

is valid; a similar result naturally holds for the functions $p_n(z)$, $\pi_n(z)$.

By Theorem I there exist polynomials $p_n(z)$ of respective degrees $n$ such that

$$|f(z) - p_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } C.$$

This same inequality is valid almost everywhere for $w$ on $\gamma$, for the two methods of defining the functions $f(z)$ and $p_n(z)$ on $\gamma$, namely by radial approach and by direct correspondence $w=\phi(z)$, $z=\psi(w)$, are equivalent almost everywhere on $\gamma$. Hence we have

$$\int_\gamma |f(z) - p_n(z)|^p n(\theta)d\theta \leqq \frac{M'}{R_1^{np}},$$

from which follows inequality (5) as before.

The proof of the theorem now follows without essential change from the proof given for Theorem IVa; the relation of the curves $C_{R_1'}$ and $C_{R_1}$ is discussed in more detail below in connection with Theorem V.

7. **Generalizations.** Theorems III and IV persist even for the case $p=1$. The proofs in that case are even more simple than the proofs we have given in detail, and the modifications are left to the reader. This is true for all of the later theorems to be proved; they are valid also for $p=1$, although the proofs are slightly different from those given for an arbitrary value of $p$ greater than unity.

There are cases of approximation analogous to Theorems II, III, and IV, where approximation with an arbitrarily small preassigned error is not possible. For instance, the function $f(z)$ may no longer be analytic interior to $C$ but may have a singularity there. Or the function $f(z)$ may be analytic as prescribed in Theorem III, but the approximating functions may (in Theorems II, III, or IV) all be required to take on values at prescribed points interior to $C$ different from the values of $f(z)$ at those points. In the simplest cases, this subject has been treated by F. Riesz,* Julia,† and the

* Acta Mathematica, vol. 42 (1919), pp. 145–171.

† Bulletin des Sciences Mathématiques, vol. 62 (1927), pp. 370–384; Annales de l'École Normale, (3), vol. 44 (1927), pp. 289–309.

present writer.* The complete extensions of Theorems III and IV still remain to be given, however, and the present writer hopes to consider them on another occasion. We remark, though, that the requiring (in addition to the other requirements) of the approximating functions to coincide with the functions approximated at a finite number of points interior to $C$ does not alter any of the results of the present paper.

8. **Surface integrals.** Thus far we have considered primarily line integrals as measures of the closeness of approximation. We now take up the possibility of using surface integrals instead. Let $C$ be an arbitrary closed Jordan region, let the function $f(z)$ be analytic in the closed region, and let the norm function $n(z)$ be positive and continuous in $C$. The measure of approximation of the polynomial $p_n(z)$ to $f(z)$ shall be†

$$\iint_C |f(z) - p_n(z)|^p \, n(z)dS, \tag{7}$$

and it follows from the method used by Julia (loc. cit.) that the polynomial $\pi_n(z)$ of best approximation exists and is unique. We now prove a result for more general regions than Jordan regions.

THEOREM V. *Let $C$ be an arbitrary limited closed region. If the function $f(z)$ is analytic interior to $C_R$, then the sequence $\{\pi_n(z)\}$ of polynomials of best approximation to $f(z)$ in the sense of least weighted pth powers over the area $C$ converges to $f(z)$ interior to $C_R$, uniformly on an arbitrary closed point set interior to $C_R$.*

It is of course true (compare §2) that the boundary of $C$ may separate the plane into more than two regions. Let $D$ denote the region which is the aggregate of all points which can be connected with the point at infinity by Jordan arcs which do not meet $C$. The curve $C_R$ is defined as usual as the transform of the circle $|w| = R$ when $D$ is mapped onto the exterior of $|w| = 1$ in such a way that the points at infinity in the $z$- and $w$-planes correspond to each other. The region $C$ of course lies interior to $C_R$. Theorem I applies here, if the point set $C$ (of Theorem I) is interpreted as the complement (with the respect to the entire plane) of $D$.

Let $R_1$ be an arbitrary positive number less than $R$. Then by Theorem I there exist polynomials $p_n(z)$ of respective degrees $n$ such that

$$|f(z) - p_n(z)| \leqq \frac{M}{R_1^n}, \quad z \text{ on } C.$$

* See the reference in the footnote in connection with Theorem II.

† In case the boundary of $C$ has area, either upper or lower integral may be used here.

Thus we have (7) less than or equal to some $M'/R_1^{np}$; the same inequality holds if $p_n(z)$ is replaced by $\pi_n(z)$, and when we notice that $n(z)$ has a positive minimum on $C$, we have

$$\iint_C |f(z) - \pi_n(z)|^p \, dS \leqq \frac{M''}{R_1^{np}}.$$

Let $K$ be an arbitrary circle of radius $r$ interior to $C$. Then we have

$$f(z) - \pi_n(z) = \frac{1}{\pi r^2} \iint_K [f(t) - \pi_n(t)] dS,$$

where $z$ denotes the center of $K$ and $t$ is the variable of integration. In fact this relation can be deduced by a single integration from Gauss's mean-value theorem for harmonic functions.* Let us apply inequality (2), setting $F(x) = |f(t) - \pi_n(t)|^p$, $\alpha = 1/p$, $G(x) = 1$. We have

$$|f(z) - \pi_n(z)| \leqq \frac{1}{(\pi r^2)^{1/p}} \left\{ \iint_K |f(t) - \pi_n(t)|^p \, dS \right\}^{1/p};$$

the integral over $K$ is not greater than the integral over $C$, so we have

$$|f(z) - \pi_n(z)| \leqq \frac{M'''}{R_1^n},$$

where $M'''$ depends on $r$ but not on $z$. This inequality holds then for every $z$, provided merely that $z$ is interior to $C$ and at a distance at least $r$ from the boundary of $C$. The inequality thus holds, for properly chosen $M'''$, for $z$ on an arbitrary Jordan curve $\Gamma'$ interior to $C$. Since $R_1 < R$ and $\Gamma'$ are both arbitrary, Theorem V follows from Theorem I and the results of Carathéodory previously used.

It will be noticed that the situation is more complicated here than previously, for $C$ is no longer a Jordan region. Nevertheless the results of Carathéodory apply, for $D$ is a region whose complete boundary lies in the (closed) region $C$, and hence when $\Gamma'$ is uniformly near the boundary of $C$, the curve $C_{R_1}'$ is uniformly near the curve $C_{R_1}$. The curves $C_{R_1}$ and $C_{R_1}'$ are the transforms of the circle $|w| = R_1$ when the region $D$ and the exterior of $\Gamma'$ respectively are mapped onto $|w| > 1$ so that the points at infinity correspond to each other. The region $D$ is the kernel of the exteriors of a properly chosen

* For the method used here, compare Kellogg, *Potential Theory*, Berlin, 1929, p. 268. The method is due to Zaremba; see Bouligand, Mémorial des Sciences Mathématiques, Fascicule 11, pp. 20–21.

sequence of curves $\Gamma'$. Carathéodory's Satz V applies directly, it being noticed that the convergence of the functions involved is uniform.*

For the particular case $p=2$, $n(z)\equiv 1$, Theorem V treats the general method of approximation used by Bergmann,† Bochner,‡ and Carleman.§ Again we have approximating polynomials of the form

$$\pi_n(z) = a_0P_0(z) + a_1P_1(z) + \cdots + a_nP_n(z),$$

where the polynomials $P_i(z)$ are independent of $f(z)$ and where the coefficients $a_i (i \leqq n)$ are independent of $n$. It will be noticed that here too we have a solution of the problem of Faber mentioned above, and that the present result holds for much more general regions than Jordan regions.

Theorem V requires more of the function $f(z)$ than that it shall be analytic in the closed region $C$, if the boundary of $C$ divides the plane into more than two distinct regions. Whatever $C$ may be, if $f(z)$ has a singularity exterior to the region $D$ used above, uniform approximation of $f(z)$ in $C$ with an arbitrarily small error is impossible; if $f(z)$ has no singularity except in $D$, then $f(z)$ is analytic interior to some $C_R$ and the hypothesis of Theorem V is satisfied. In particular, *if $C$ is an arbitrary limited region whose boundary divides the plane into but two distinct regions, a function $f(z)$ analytic in the (closed) region $C$ is analytic within some curve $C_R$, and hence the sequence $\{\pi_n(z)\}$ of polynomials of best approximation to $f(z)$ on $C$ converges uniformly in the closed region $C$*, in fact converges interior to $C_R$, uniformly on any closed point set interior to $C_R$. This same remark applies in the case of Theorem IV, if $C$ is simply connected; compare the footnote to Theorem XII.

We do not carry out the analogue of Theorem V after mapping of $C$ onto the unit circle $\gamma$; this is easy to do, but due to the arbitrariness of the function $n(z)$, would yield little new. A double integral over $C$ is equal to the corresponding double integral over the interior of $\gamma$, with an additional factor in the integrand, the square of the derivative of the mapping function. It may be remarked that the function $n(z)$ need not be continuous, although for simplicity we have made that assumption, but even in Theorems III and IV as well need merely be positive, integrable, and bounded from zero. We do not, however, mean to imply the complete equivalence of Theorem V and the analogous theorem obtained by studying approximation measured

* The theorem for harmonic functions (used below) corresponding to Theorem I is not an exact analogue of Theorem I, but is to be found as Theorem III of a paper by the present writer, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 591–598.

† Mathematische Annalen, vol. 86 (1922), pp. 238–271.

‡ Mathematische Zeitschrift, vol. 14 (1922), pp. 180–207.

§ Arkiv för Matematik, Astronomi, och Fysik, vol. 17 (1922–23).

over the area interior to $\gamma$. The latter can be obtained with ease by the reader, in the special case corresponding to Theorem V that $C$ is simply connected.

9. **Approximation to harmonic functions.** We turn now to the analogues of Theorems I-V, for approximation to harmonic functions by harmonic polynomials. The results and methods are in the main so similar to those already used that we need not enter into full detail. The analogue of Theorem I is not difficult to prove,* provided that $C$ is a Jordan region, and leads† directly to the analogue of Theorem II, also for the case that $C$ is a Jordan region, and even if $C$ is an arbitrary limited region. This result, the analogue of Theorem II, naturally persists even though we consider the harmonic Tchebycheff polynomial with the norm function $n(x, y)$.

The analogue of Theorem IV is likewise easy to prove:

THEOREM VI. *Let the function $u(x, y)$ be harmonic within $C_R$, where $C$ is an arbitrary closed limited simply connected region. Then the sequence of harmonic polynomials $\{\pi_n(x, y)\}$ of best approximation to $u(x, y)$ on $C$ in the sense of least weighted pth powers as measured on $\gamma$ converges to $u(x, y)$ throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

In the proof of Theorem VI it is essential to know, in order to use Poisson's integral, that for such a function as $u(x, y)$ the functional values on $\gamma$ found by conformal transformation coincide almost everywhere with the functional values found by radial approach to $\gamma$ after conformal transformation. This presents no difficulty, since $\lim_{r\to 1,\, r<1} \psi(re^{i\theta})$ exists for almost all values of $\theta$; such a value $\theta'$ corresponds to an accessible point of the boundary of $C$, and the radius $\theta=\theta'$ of $\gamma$ corresponds to a Jordan arc interior to $C$ abutting in $\psi(e^{i\theta'})$. The limit of $u(x, y)$ along this arc is naturally the value of $u(x, y)$ at the end point.

In the proof of Theorem VI there enters the derivative of Green's function for the circle $\gamma$ in the place of the function $1/(t-w)$ of Theorem IV.

This general situation, of considering best approximation to a harmonic function by harmonic polynomials in the sense of least $p$th powers as measured on $\gamma$, has recently been studied by Julia (loc. cit.), and important general properties derived.

The analogue of Theorem III is not so simple, for the precise analogue of Cauchy's integral is Green's formula

---

* See Walsh, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 591–598.

† Walsh, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 499–544; especially p. 513. The theorem is there stated and proved only for a Jordan region $C$, but the proof can be given in the more general case. See Theorem III of the paper referred to in the preceding footnote.

$$(8) \qquad u(x, y) - \pi_n(x, y) = \frac{1}{2\pi}\int_\Gamma [u(x, y) - \pi_n(x, y)]\frac{\partial G}{\partial n}ds,$$

where $G$ is Green's function for the region, and comparatively little seems to be known about $\partial G/\partial n$ for an arbitrary rectifiable Jordan curve $\Gamma$. We may take an extreme case, however, and assume the curve $\Gamma$ to have continuous curvature. It follows that $\partial G/\partial n$ is continuous on $\Gamma$, and we may state

THEOREM VII. *Let the function $u(x, y)$ be harmonic within $C_R$, where the point set $C$ is the interior of the Jordan curve $\Gamma$, assumed to have continuous curvature. Then the sequence of harmonic polynomials $\{\pi_n(x, y)\}$ of best approximation to $u(x, y)$ on $C$ in the sense of least weighted pth powers as measured on $\Gamma$ converges to $u(x, y)$ throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$.*

This particular measure of approximation for $p=2$, $n(s)\equiv 1$, has been considered by S. Bernstein,* Brillouin,† and Picone,‡ and for this restricted case Theorem VII has been proved by Merriman§ by the present methods.

The analogue of Theorem V presents not the slightest difficulty:

THEOREM VIII. *Let $C$ be an arbitrary limited closed region. If the function $u(x, y)$ is harmonic interior to $C_R$, then the sequence $\{\pi_n(x, y)\}$ of harmonic polynomials of best approximation to $u(x, y)$ in the sense of least weighted pth powers over the area $C$ converges to $u(x, y)$ interior to $C_R$, uniformly on an arbitrary closed point set interior to $C_R$.*

Here too we have for the case $p=2$, $n(x, y)\equiv 1$, approximating polynomials $\pi_n(x, y)$ of the form

$$\pi_n(x, y) = a_0P_0(x, y) + a_1P_1(x, y) + \cdots + a_nP_n(x, y),$$

where the polynomials $P_i(x, y)$ are independent of $u(x, y)$ and where the coefficients $a_i (i \leqq n)$ are independent of $n$. In particular *if $C$ is an arbitrary limited region whose boundary divides the plane into but two distinct regions, a function $u(x, y)$ harmonic in (the closed region) $C$ is harmonic within some curve $C_R$, and hence the sequence $\{\pi_n(x, y)\}$ of polynomials of best approximation to $u(x, y)$ on $C$ converges uniformly to $u(x, y)$ in the closed region $C$,* in fact converges interior to $C_R$, uniformly on any closed point set interior to $C_R$.

* Paris Comptes Rendus, vol. 148 (1909), pp. 1306–1308.

† Annales de Physique, vol. 6 (1916), pp. 137–223.

‡ Rendiconti dei Lincei, 1922, pp. 357–359. These three references were unintentionally omitted in the writer's report referred to above, on approximation of harmonic functions by harmonic polynomials.

§ *On the expansion of harmonic functions in terms of normal orthogonal harmonic polynomials*, not yet published. Presented to the American Mathematical Society in April, 1928.

10. **Generalizations.** We make a few remarks about the existence and uniqueness of the polynomials of best approximation. The harmonic Tchebycheff polynomial of best approximation with a norm function unity to a function harmonic interior to an arbitrary Jordan curve $\Gamma$, continuous in the corresponding closed region, has been considered independently by Julia, in a paper presented to the Bologna Congress (September 1928) and not yet published, and by the present writer.* It is of course immaterial here whether we study approximation to $u(x, y)$ on $\Gamma$ or in the closed region $C$ bounded by $\Gamma$. The Tchebycheff polynomial always exists in this case, but is not necessarily unique; this same remark holds for an arbitrary limited region $C$ bounded by a point set $\Gamma$, and whether or not the norm function $n(x, y)$ (continuous on $C$) is unity.

In any of the situations of Theorems V–VIII, the harmonic polynomial of best approximation always exists, for in every case boundedness for a sequence of approximating polynomials of the measure of approximation in $C$ implies uniform boundedness of the sequence of polynomials in an arbitrary closed region interior to $C$, and the method as given by Julia applies. In these same situations, the harmonic polynomial of best approximation is likewise unique; this also follows from the method given by Julia for the case of Theorem VI, $n(s) \equiv 1$, in the paper to which reference has been made several times. The method and result apply too in the cases of approximation to analytic functions already taken up, where approximation is measured by an integral. The same method and result apply also for approximation to analytic and harmonic functions in the similar cases to be treated below. The uniqueness of the polynomials of best approximation, it may be pointed out, does not enter in any way into Theorems II–VIII.

It will be noticed that the theorems we have stated for harmonic functions, like those for analytic functions, are not the most general that we might have stated. They involve, in fact, not necessarily the sequence of polynomials of *best* approximation, but any sequence of polynomials which converges, according to some measure of convergence, like a convergent geometric series. Let us for definiteness describe the situation as in Theorems III–V for approximation to an analytic function, although of course the situation is similar for harmonic functions. We can show (from Theorem I) under certain restrictions on the region $C$, that if $f(z)$ is analytic interior to $C_R$, there exists a sequence of polynomials of respective degrees $n$ such that the measure of approximation (considered in any one of several ways) is of the order $1/R_1^{np}$ for an arbitrary $R_1 < R$. We can show on the other hand

* Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 499–544; §4.

that any sequence of polynomials of respective degrees $n$ such that the measure of approximation is of order $1/R_1^{np}$ for an arbitrary $R_1<R$, converges to $f(z)$ not merely on $C$, but throughout the interior of $C_R$, uniformly on any closed point set interior to $C_R$. This conclusion naturally follows, then, for the polynomials of best approximation.

A similar remark applies to the discussion of §§11 and 12 below.

It will be noticed too that each of Theorems II–V gives a result relative to the degree of the best approximation to the given function. In Theorem II, for instance, suppose $f(z)$ actually to have a singularity on $C_R$. If $R_1$ is an arbitrary number less than $R$, we surely have for some $M_1$

$$|f(z)-\pi_n(z)| \leqq M_1/R_1^n, \quad z \text{ on } C.$$

Yet there exists no number $R_2$ greater than $R$ such that

$$|f(z)-\pi_n(z)| \leqq M_2/R_2^n, \quad z \text{ on } C.$$

Similar results for the other measures of approximation can obviously be written down, for analytic functions and for harmonic functions.

11. **Approximation to less restricted analytic functions.** It is worth noticing that the inequalities we have been using in proving the overconvergence of the sequences of polynomials are precisely the natural equalities that might be used in proving the simple convergence of the approximating polynomials when less stringent requirements are put on the functions approximated. In every case corresponding to Theorems III–VIII where we deal with a Jordan region $C$, if the function to be approximated is analytic or harmonic in the interior of the region $C$ and continuous in the closed region, then the sequence of polynomials of best approximation converges to the function interior to $C$, uniformly on any closed point set interior to $C$. Let us prove this fact, for instance, in the situation corresponding to Theorem III.*

THEOREM IX. *Let the function $f(z)$ be analytic interior to $C$, continuous in the corresponding closed region, where $C$ is the interior of the rectifiable Jordan curve $\Gamma$. Then the sequence of polynomials $\{\pi_n(z)\}$ of best approximation to $f(z)$ on $C$ in the sense of weighted pth powers as measured on $\Gamma$ converges to $f(z)$ throughout the interior of $C$, uniformly on any closed point set interior to $C$.*

We introduce the following notation for the *measure of approximation*:

* For the sequence of Tchebycheff polynomials, we have convergence to the function approximated, uniformly in the *closed* region $C$. Compare, for more general regions $C$, the end of §2 and the beginning of §12.

$$\epsilon_n = \int_\Gamma | f(z) - \pi_n(z) |^p \, n(s)ds.$$

The function $f(z)$ can be uniformly approximated by a polynomial as closely as desired in the closed region $C$, so $\lim_{n\to\infty} \epsilon_n = 0$. We have, as in (4),

$$\int_\Gamma | f(z) - \pi_n(z) |^p \, ds \leqq \int_\Gamma | f(z) - \pi_n(z) |^p \frac{n(s)}{n'} ds \leqq \frac{\epsilon_n}{n'}.$$

We apply Cauchy's integral formula, as to (4), and find

$$| f(z) - \pi_n(z) | \leqq N\epsilon_n,$$

where $z$ is restricted to lie on an arbitrary closed point set interior to $C$; the number $N$ depends on this point set but not on $n$. Theorem IX follows at once.

We state explicitly the theorems corresponding to Theorems IV–VIII as well. These theorems are for the most part not entirely new; their present interest lies in their proof by the methods already used. Many of these methods are used here for the first time to prove these particular theorems. The theorems are, however, more general than those proved elsewhere, either with respect to measure of approximation, by introducing a weight function and considering an arbitrary value of $p$ instead of $p=2$, or in treating regions more general than those treated hitherto. References to the literature are to be found in connection with the theorems previously given.

THEOREM Xa. *Let the function $f(z)$ be analytic interior to $C$, continuous in the corresponding closed region, where $C$ is the interior of an arbitrary Jordan curve $\Gamma$. Then the sequence of polynomials $\{\pi_n(z)\}$ of best approximation to $f(z)$ on $C$ in the sense of least weighted pth powers as measured on the circumference $\gamma$ converges to $f(z)$ interior to $C$, uniformly on any closed point set interior to $C$.*

THEOREM X. *Let $C$ be an arbitrary limited simply connected region. If the function $f(z)$ is analytic interior to $C$, continuous in the corresponding closed region, and if $f(z)$ can be approximated in the closed region as closely as desired by a polynomial in $z$, then the sequence $\{\pi_n(z)\}$ of polynomials of best approximation to $f(z)$ in the sense of weighted pth powers as measured on the circumference $\gamma$ converges to $f(z)$ interior to $C$, uniformly on any closed point set interior to $C$.*

THEOREM XI. *Let $C$ be an arbitrary Jordan region. If the function $f(z)$ is analytic interior to $C$, continuous in the corresponding closed region, then the sequence $\{\pi_n(z)\}$ of polynomials of best approximation to $f(z)$ in the sense of weighted pth powers over the area $C$ converges to $f(z)$ interior to $C$, uniformly on any closed point set interior to $C$.*

In Theorems Xa and XI the region $C$ is restricted to be a Jordan region, for it is only for such a region that we know that an arbitrary function $f(z)$ analytic in the interior and continuous in the closed region can be approximated as closely as desired by a polynomial in $z$. The formal requirements for approximation exist in Theorem XI, however, if $C$ is an arbitrary limited region, simply connected or not, and in particular if the weight function $n(z)$ is chosen as unity, then for $p=2$ we have a set of polynomials $\{P_n(z)\}$ *belonging to* $C$. Any function $f(z)$ which can be approximated arbitrarily closely leads in Theorems X, Xa, and XI to a series

$$f(z) = a_0P_0(z) + a_1P_1(z) + \cdots$$

where the polynomials $P_i(z)$ depend on $C$ but not on $f(z)$, where the coefficients $a_i$ depend only on $f(z)$ and where the series converges to the sum $f(z)$ interior to $C$, uniformly on an arbitrary point set interior to $C$. The partial sums of the series are the polynomials of best approximation.

If $C$ is a circle, the polynomials $P_n(z)$ are, except for numerical factors, the polynomials $z^n$ respectively.

This remark on the special form of the polynomials $\pi_n(z)$ applies also in connection with

THEOREM XII. *Let $C$ be an arbitrary limited region. If the function $f(z)$ is analytic interior to $C$, continuous in the corresponding closed region, and if $f(z)$ can be approximated in the closed region as closely as desired* by a polynomial in $z$, then the sequence $\{\pi_n(z)\}$ of polynomials of best approximation to $f(z)$ in the sense of weighted $p$th powers over the area $C$ converges to $f(z)$ interior to $C$, uniformly on any closed point set interior to $C$.*

Theorem XII has the advantage over Theorem X of holding for multiply connected regions.

We refrain from writing down the analogues of Theorems XI and XII obtained by conformal mapping, considering approximation in the sense of

* It may be noticed that here and elsewhere in the present paper this measure of approximation may be taken either in the sense of Tchebycheff or in the sense of weighted $p$th powers as used in the theorem itself. We remark that analyticity of $f(z)$ in the closed region $C$ does not ensure the satisfaction of the condition of the theorem, except in the case that the boundary of $C$ divides the plane into precisely two regions. Compare Walsh, Mathematische Annalen, vol. 96 (1926), pp. 430–436; 437–450, especially p. 433 and p. 441, footnote. In particular, if $C$ is not simply connected, the function $f(z)$ can have no singularities interior to a Jordan curve $K$ which lies entirely interior to $C$. For the sequence $\{\pi_n(z)\}$ converges uniformly to $f(z)$ on $K$, hence represents a function analytic everywhere interior to $K$ coinciding with $f(z)$ on $K$ and therefore coinciding with $f(z)$ everywhere interior to $K$.

least weighted $p$th powers over the area interior to $\gamma$ in case $C$ is simply connected. Compare the remark made relative to Theorem V.

12. **Approximation to less restricted harmonic functions.** The theorems on harmonic functions are also not difficult to treat.

We consider first the Tchebycheff measure of approximation. Let $C'$ be an arbitrary closed limited region and let $C$ denote the totality of closed limited regions into which the boundary of $C'$ divides the plane. The boundary points of $C$ (considered as a set of individual regions, not as a point set) are precisely the boundary points of $C'$. If the function $u(x, y)$ is harmonic interior to $C'$, continuous in the closed region, it is immaterial whether we approximate $u(x, y)$ in $C'$, or on the boundary of $C$ or $C'$. On all of these point sets, the maximum of

$$| u(x, y) - p_n(x, y) |,$$

where $p_n(x, y)$ is a harmonic polynomial, is one and the same. This fact also holds for the approximation of $u(x, y)$ on $C$, if by $u(x, y)$ we mean not necessarily the harmonic extension of the given function $u(x, y)$ defined in $C'$, but the function $u(x, y)$ defined on the entire point set $C$ as the function which is continuous on $C$, which in the boundary points of $C$ (i.e. considered as a set of regions, not as a point set) takes on the values of the given function $u(x, y)$, and which is harmonic in the interior points of $C$.* If the sequence of Tchebycheff harmonic polynomials, or any other set of harmonic polynomials, converges to $u(x, y)$ on the boundary of $C'$, that sequence converges on the entire point set $C$ to the function $u(x, y)$ as just defined. Only such a function $u(x, y)$ actually or potentially defined on $C$ can be uniformly approximated in the closed region $C'$ by harmonic polynomials. A necessary and sufficient condition that such a function can be so approximated, which is a necessary and sufficient condition for the convergence of the sequence of Tchebycheff polynomials to the function, is contained in the second paper just referred to.

THEOREM XIII. *Let the function $u(x, y)$ be harmonic interior to $C$, continuous in the corresponding closed region, where $C$ is the interior of an arbitrary Jordan curve $\Gamma$ with continuous curvature. Then the sequence of harmonic polynomials $\{\pi_n(x, y)\}$ of best approximation to $u(x, y)$ on $C$ in the sense of weighted pth powers as measured on $\Gamma$, converges to $u(x, y)$ throughout the interior of $C$, uniformly on any closed point set interior to $C$.*

* Compare Walsh, Bulletin of the American Mathematical Society, vol. 33 (1927), pp. 591–598; p. 597.
Also Crelle's Journal, vol. 159 (1928), pp. 197–209; pp. 204–205.

THEOREM XIV. *Let $C$ be an arbitrary limited simply connected region whose boundary is also the boundary of an infinite region. If the function $u(x, y)$ is harmonic interior to $C$, continuous in the closed region, then the sequence $\{\pi_n(x, y)\}$ of harmonic polynomials of best approximation to $u(x, y)$ on $C$ in the sense of weighted pth powers as measured on the circumference $\gamma$ converges to $u(x, y)$ interior to $C$, uniformly on any closed point set interior to $C$.**

In particular if the weight function is unity and $p=2$, we have a set of harmonic polynomials $\{P_n(x, y)\}$ belonging to the region $C$, and the sequence $\{\pi_n(x, y)\}$ corresponds to the partial sums of the series

$$u(x, y) = a_0P_0(x, y) + a_1P_1(x, y) + \cdots,$$

where the coefficients $a_i$ depend only on $u(x, y)$. This remark can be extended to the most general limited region.

THEOREM XV. *Let $C$ be an arbitrary limited simply connected region. If the function $u(x, y)$ is harmonic interior to $C$, continuous in the corresponding closed region, and if $u(x, y)$ can be uniformly approximated in the closed region as closely as desired by a harmonic polynomial, then the sequence $\{\pi_n(x, y)\}$ of harmonic polynomials of best approximation to $u(x, y)$ on $C$ in the sense of weighted pth powers as measured on the circumference $\gamma$ converges to $u(x, y)$ interior to $C$, uniformly on any closed point set interior to $C$.*

THEOREM XVI. *Let $C$ be an arbitrary limited region whose boundary is also the boundary of an infinite region. If the function $u(x, y)$ is harmonic interior to $C$, continuous in the closed region, then the sequence $\{\pi_n(x, y)\}$ of harmonic polynomials of best approximation to $u(x, y)$ on $C$ in the sense of weighted pth powers as measured over the area $C$ converges to $u(x, y)$ interior to $C$, uniformly on any closed point set interior to $C$.*

If $p=2$ and the weight function is unity, we have a set $\{P_n(x, y)\}$ of harmonic polynomials belonging to the region $C$, and the sequence $\{\pi_n(x, y)\}$ corresponds to the partial sums of the series

$$u(x,y) = a_0P_0(x,y) + a_1P_1(x,y) + \cdots,$$

where the coefficients $a_i$ depend only on $u(x, y)$ and where the polynomials $P_n(x, y)$ are normal and orthogonal with respect to the area $C$. When $C$ is

* Here we apply a general theorem on approximation to harmonic functions; see the paper by the present writer in Crelle's Journal to which reference has just been made. That paper was presented to the American Mathematical Society in December, 1926, and an abstract of it was published in the Bulletin of that Society, vol. 33 (1927), p. 150. Compare also Julia, a paper presented to the Bologna Congress, September, 1928, and not yet published.

a circle, the polynomials $P_n(x, y)$ are precisely those which lead to the expansion of $u(x, y)$ in Fourier's series. In this case, the expansion of $u(x, y)$ is the same whether we have best approximation in the sense of least squares over the area $C$ or over the circumference of $C$.

THEOREM XVII. *Let $C$ be an arbitrary limited region. If the function $u(x, y)$ is harmonic interior to $C$, continuous in the closed region, and if $u(x, y)$ can be approximated in the closed region as closely as desired by a harmonic polynomial,* then the sequence $\{\pi_n(x, y)\}$ of harmonic polynomials of best approximation to $u(x, y)$ on $C$ in the sense of weighted $p$th powers as measured over the area $C$ converges to $u(x, y)$ interior to $C$, uniformly on any closed point set interior to $C$.*

Theorem XVII holds even for harmonic functions in $n$ dimensions. In particular if $C$ is bounded by a simple closed surface, and if $u(x, y, \cdots)$ is harmonic in the closed region, the hypothesis of the theorem is satisfied.

Nowhere in the suggested proofs of Theorems IX–XVII is there used the fact that the sequence $\{\pi_n(z)\}$ or $\{\pi_n(x, y)\}$ is the sequence of polynomials of *best* approximation, although the existence and uniqueness of those polynomials of best approximation are assured as in connection with the previous theorems. It is sufficient for the conclusion of Theorems IX–XVII in each case if $\epsilon_n$, the measure of approximation of the polynomial to the function (compare the proof of Theorem IX), approaches zero with $1/n$. Moreover, we have assumed for simplicity that the functions $f(z)$ and $u(x, y)$ are continuous in the closed regions considered, but that restriction can be materially lightened in the several cases, as an inspection of the proofs will show.

13. **Application to the Dirichlet problem.** As Julia points out, the fact that for a Jordan region the harmonic polynomial of best approximation to a harmonic function, either in the sense of Tchebycheff or in the sense of weighted $p$th powers as measured on $\gamma$ or on $\Gamma$, refers merely to the boundary of the region, gives us a theoretical solution of the Dirichlet problem. The Tchebycheff harmonic polynomials of best approximation are the same for the boundary values and for the harmonic function taking on those boundary values (i.e. the corresponding solution of the Dirichlet problem), and the sequence of

* It will be noticed in particular that if $C$ is not simply connected, the function $u(x, y)$ and its harmonic extension can have no singularities interior to an arbitrary Jordan curve $K$ which lies entirely interior to $C$. For the sequence $\{\pi_n(x, y)\}$ converges uniformly to the limit $u(x,y)$ on $K$, hence represents a function harmonic everywhere interior to $K$, which coincides at some points interior to $K$ (namely at points interior to both $K$ and $C$) with $u(x, y)$, and which therefore coincides with $u(x, y)$ everywhere interior to $K$.

those polynomials or of the other polynomials mentioned converges interior to the given region to the solution of the Dirichlet problem which corresponds to the given boundary values. With the Tchebycheff method of approximation the actual determination of the polynomials of best approximation seems to be extremely difficult. In the cases of Theorems XIII–XV the formulas for the polynomials of best approximation ($p=2$, $n(s)\equiv 1$) involve only comparatively simple line integrals, so the method of solving the Dirichlet problem has some practical value. The numerical computation can in certain cases actually be carried out, as has been done notably by Bergmann.*

The question of the actual convergence on $\Gamma$ of the approximating sequences that we have been considering in §§11 and 12 is an interesting one.† If approximation is in the sense of Tchebycheff, the sequence of polynomials of best approximation naturally converges uniformly in the closed region $C$ in every case we have treated. If approximation is measured by a line integral, in every case the measure $\epsilon_n$ of approximation approaches zero with $1/n$, and it follows that we have

$$\operatorname*{qlim}_{n\to\infty} \pi_n(z) = f(z), \quad \operatorname*{qlim}_{n\to\infty} \pi_n(x, y) = u(x, y),$$

in the sense of quasiconvergence‡ on $\Gamma$. That is to say, from every subsequence of the $\pi_n(z)$ [or $\pi_n(x, y)$] can be extracted a new subsequence which converges almost everywhere on $\Gamma$ to $f(z)$ [or $u(x, y)$].

If $\epsilon_n$ is an integral taken over $\Gamma$, measure (in the term quasiconvergence) is to be taken in the sense of arc length on $\Gamma$. If $\epsilon_n$ is an integral over $\gamma$, we have quasiconvergence on $\gamma$ and hence on $\Gamma$, but measure is here to be taken in the sense of arc length on $\gamma$. Nevertheless, even in this case, measure may be taken in the sense of arc length on $\Gamma$ if $\Gamma$ is rectifiable.

---

* See for instance Zeitschrift für Angewandte Mathematik und Mechanik, vol. 8 (1928), pp. 402–413.

† There is also the corresponding question in §§1–10 of the convergence on $C_R$; this question seems to be entirely unanswered in the general case. S. Bernstein has established some results in this connection for approximation on a point set $C$ which is an interval of the axis of reals.

‡ See Walsh, these Transactions, vol. 32 (1930), pp. 335–390.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# ON BIRKHOFF'S PFAFFIAN SYSTEMS*

BY
LUCIEN FERAUD

## INTRODUCTION

The *Pfaffian systems* to which this paper is devoted were defined and considered for the first time in 1920 by G. D. Birkhoff in the Chicago Colloquium lectures. Later he was led to a consideration of them in his memoir† *Stability and the equations of dynamics* and published in his book‡ *Dynamical Systems* the demonstration of the important results he announced there.

I shall set down later the precise definition of those systems; for the present it is sufficient to say that the Hamiltonian form can be obtained from them by a natural particularization, and to give their fundamental property: namely, of being of the same form after any point transformation of the dependent variables. This last statement must be compared with the well known invariance of the Hamiltonian form under the canonical transformations, and so one is naturally led to the question of the analytic significance of the Pfaffian form, and that is the first point studied here. First of all, any Pfaffian system is related to a differential expression $\omega$ so that the choice of a special form for the system is equivalent to the choice of $\omega$. Moreover it is easy to see that any system under consideration admits $\int\omega$ as a relative integral invariant and is precisely the "*characteristic system*" of $\int\omega'$. So without referring to any dynamical problem, I first intend to make as clear as possible the interest attached to the Pfaffian form and the necessity of its use. For the same purpose in the case where a certain number of integrals are known, I give an extension of Poisson's and Lie's theorems and also several other properties relative to the use of the known integrals for reducing the order of the system. At the end of the first section I find a very simple expression for the integral invariant of order equal to the order of the system, and consequently a noteworthy form for the "*last multiplier*"; this last multiplier is the square root of the determinant formed with all the coefficients appearing in the first members in the Pfaffian system.

As an elucidation of the way in which one can be led to the consideration of Pfaffian systems in dynamical problems, I study in the second section the

* Presented to the Society, September 11, 1930; received by the editors June 29, 1929.
† American Journal of Mathematics, vol. 49 (1927), p. 1.
‡ American Mathematical Society Colloquium Publications, vol. IX, 1927.

reduction of the Lagrangian to the Pfaffian form. The entire question depends on a system of simultaneous partial differential equations in the $F_i$, $G_i$ (which define a point change of variables) and the $X_i$, $Z$ of the form $\omega$ (which determines the Pfaffian system under consideration). Hence a whole class of problems arises when we regard some of the functions $F_i$, $G_i$, $X_i$, $Z$ as known and the others as a solution of the system of simultaneous partial differential equations. In applying this general process I obtain first a reduction to the normal form in a way which presents the advantage of leading directly to the result, and I give afterwards a reduction to the Hamiltonian form.

At the beginning of the third section I briefly review, following there Birkhoff's book quoted above, the most important properties of the Pfaffian systems considered from a dynamical point of view, together with the different problems to which they can be applied. Then, in applying results of the first section, I show how, after the greatest possible reduction, the final order of the dynamical system for the three-body and some similar problems can be obtained in a very simple way.

Finally I consider the "prepared" Pfaffian expression $\omega$ which Birkhoff attaches to a Pfaffian system in the vicinity of an equilibrium point, and I study the reduction of $\omega$ to a canonical form $\bar{\omega}$, in fact the reduction of $\omega$ to the sum of $\omega$ and of a perfect differential $dE$.

This is a problem of fundamental importance from a dynamical point of view, of which Birkhoff* has given a formal solution by a process that does not introduce small divisors. This last fact suggests at once the possibility of establishing the considered reduction by the use of convergent series, but it does not give a demonstration of that possibility. So I was led to consider this problem not only in a formal sense but in an actual sense. For this problem I have obtained a solution which I have shown to be both formal and actual, in the case of two degrees of freedom. Indeed at the end of this paper I demonstrate, when $m=2$, that the considered reduction can always be established by means of analytic functions.†

## I

We shall consider systems of differential equations of the following form:

$$\text{(I)} \qquad \sum_{j=1}^{2m} a_{ij}\frac{dX_j}{dt} + \frac{\partial X_i}{\partial t} - \frac{\partial Z}{\partial x_i} = 0 \qquad (i = 1, 2, \cdots, 2m),$$

* American Journal of Mathematics, loc. cit.

† Some of the results of this paper have been briefly indicated in two notes published in the Paris Comptes Rendus, vol. 188, p. 1029 and p. 1144.

where $X_i$ and $Z$ are $2m+1$ functions of the $x_i$ and of $t$, and

$$a_{ij} = \frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i}.$$

Such systems give the extremals in the variational problem that can be stated by

$$\delta \int_{t_0}^{t_1} \left[ \sum_{j=1}^{2m} X_j \frac{dx_j}{dt} + Z \right] dt = 0.$$

Let us introduce now the system

$$\text{(II)} \qquad \sum_{j=1}^{2m} a_{ij} dx_j + \left( \frac{\partial X_i}{\partial t} - \frac{\partial Z}{\partial x_i} \right) dt = 0 \quad (i = 1, 2, \cdots, 2m)$$

from which the former arises when the variables $x_i$ have to be functions of $t$. The system (II) is very closely connected with the Pfaffian expression

$$\omega = \sum_{j=1}^{2m} X_j dx_j + Z dt,$$

and it is interesting to consider with some attention those connections in order to realize fully the analytic significance of systems (I). For that purpose one can refer to E. Cartan, *Leçons sur les Invariants Intégraux*,* where those connections are to be found among other still more powerful results. Here we are going to enunciate only the properties that seem to be the most important in dynamics and to recall at the same time some definitions that may facilitate a reference to the book cited.

For the system (II) we have a relative integral invariant $\int\omega$ and it is a linear integral invariant. Furthermore to every form $\omega$ there corresponds in general a system (II) and only one, admitting $\int\omega$ as a relative integral invariant. It is called the *characteristic system* of $\int\omega$ or the *associated system* of the form $\omega'$ ($\omega'$ being the bilinear covariant of the form $\omega$).† Hence the systems (II) can be defined as the systems admitting a linear relative integral invariant. If we now suppose the existence of known integrals, we can extend to our systems Poisson's theorem and also Lie's theorem‡ (evidently with the particular case $n=m$ due to Liouville) that gives the reduction of

* Paris, Librairie Hermann, 1922. See particularly chapter XII entitled *Systems admitting a relative integral-invariant.*

† It is the "first Pfaffian system" of Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge, 1927, p. 307, and also, according to Morera, the "associate of the Pfaffian expression $\omega$," Rendiconti, Accademia dei Lincei, vol. 12 (1st sem. 1903), pp. 113–122.

‡ See Levi-Civita e Amaldi, *Lezioni di Meccanica Razionale*, vol. II, part 2, p. 380; Bologna, 1927.

the order of a system when there are given $n$ integrals independent and moreover in involution.

In a more general manner Cartan gives the largest reduction in the order of the system that can be made, when $n$ integrals are known, independent but not necessarily in involution. He shows the way in which this reduction can be obtained and states a theorem giving the number by which the order of the system can be diminished. In particular if $\omega$ can be written as an exact differential with respect to the known integrals, the problem can be solved by one quadrature. Finally let us point out that whatever may be the number of integrals given, the order of the system is to be reduced by an even number, so that if we know $2k+1$ integrals, the order of the system will be reduced by at least $2k+2$. It is clear that those properties belong also to the systems (I), if none of the integrals we have to consider is ever $t=C$.

In chapter I of the same book, Cartan shows that the Hamiltonian equations can be characterized as those which admit the integral invariant $\int\omega_s$, with $\omega_s=\sum p_i\delta q_i-H\delta t$, and he calls this last expression "tenseur quantité de mouvement-énergie." By any change of variables $\omega_s$ will be transformed into a form $\omega$ and the Hamiltonian system into a system (I); furthermore, the system (I) will be determined by $\omega$, just as the Hamiltonian system is the only one that belongs to $\omega_s$. So we can conclude that in their more general form the dynamical systems can be represented by a linear differential expression $\omega$.

It is evident now that every particularization of the form $\omega$ will give a particularization for the corresponding system: the Hamiltonian system is to be obtained when $\omega$ is taken in the particular form $\omega_s$. Naturally the question arises whether a different particularization of $\omega$ can lead to a system more appropriate to the study of certain problems of dynamics and also whether avoiding the reduction to the Hamiltonian form would not be in some cases a simplification. Birkhoff* opened the way in that direction when he first defined as a *Pfaffian system*† the system (I) corresponding to a form

* Loc. cit.

† To avoid all difficulties let us make the following remarks about the different denominations that are to be used:

(1) Equations in which the first members are linear differential expressions, when not distinguishing any of the independent variables, are known as Pfaff's equations and systems of such equations as *Pfaff's systems*.

(2) The systems studied in this paper were named Pfaffian systems by G. D. Birkhoft; when there might be some confusion they can be designated as "dynamical Pfaffian systems."

(3) In a preceding note, we have already referred to the *first Pfaff's system of a form* $\omega$, so called by Whittaker.

(4) Finally, in the theory of skew-symmetric determinants we have later to deal with expressions spoken of by Cayley as "Pfaffians."

$\omega$ in which $X_i$ and $Z$ do not involve $t$. Further, he uses the name *generalized Pfaffian system* when $X_i$ and $Z$ are periodic in $t$ with the same period.

We will limit ourselves hereafter in this paper to the consideration of the dynamical Pfaffian systems. For those systems $Z=C$ is an integral, and by use of it and by elimination of $dt$, it is easy to show that any such system of order $2m$ is in fact equivalent to a system of differential equations of order $2m-2$. After this reduction the system obtained is analogous to a Pfaffian system, with the difference, however, that the functions $X_i$ and $Z$ involve $t$; we can say, in other words, that $X_i$ and $Z$ involve $t$ as in a generalized Pfaffian system but are no longer periodic in that variable.

Furthermore, the Pfaffian form is preserved by any of the following changes of variables:

$$\bar{x}_i' = f_i(x_1, x_2, \cdots, x_{2m}) \qquad (i=1, 2, \cdots, 2m),$$
$$\bar{t} = \phi(t),$$

$f_i$ and $\phi$ being $2m+1$ arbitrary functions. Let us suppose now that we have $2k+1$ integrals not involving $t$, independent of each other and independent of $Z=C$:

$$F_1(x_1, x_2, \cdots, x_{2m}) = C_1,$$
$$F_2(x_1, x_2, \cdots, x_{2m}) = C_2,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$F_{2k+1}(x_1, x_2, \cdots, x_{2m}) = C_{2k+1}.$$

The change of variables

$$\bar{x}_1 = F_1, \qquad \bar{x}_2 = F_2, \quad \cdots, \quad \bar{x}_{2k+1} = F_{2k+1},$$
$$\bar{x}_{2k+2} = x_{2k+2}, \quad \bar{x}_{2k+3} = x_{2k+3}, \cdots, \quad \bar{x}_{2m} = x_{2m}$$

transforms $\omega$ into another form $\bar{\omega}$ and $Z=C$ into another integral $\overline{Z}=C$.

The given system in which we know the $2k+1$ integrals is the system that gives the extremals in the variational problem $\delta\int\bar{\omega}=0$ when we have replaced, in $\bar{\omega}$, $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{2k+1}$ by arbitrary constants. If we write this system, we see at once that it can admit solutions only if a certain determinant (characteristic) is 0. This last equality allows us to express one of the variables $\bar{x}_{2k+2}, \cdots, \bar{x}_{2m}$ in terms of the others, and to reduce $\omega$ to a form containing only $2m-2k-2$ variables. The system which gives the extremals will be of that order $2m-2k-2$ and still of the same form, i.e., it will admit an integral such as $Z=C$, so, in dealing with this last fact and by elimination of $dt$, its order can be possibly reduced by two units more.

It is clear that the system we consider now in which $Z=C$ is an integral admits as a relative invariant integral

$$\int \sum_{i=1}^{2m} X_i dx_i.$$

From this can be obtained at once by Stokes' theorem the absolute integral invariant

$$\iint \sum_{i=1}^{2m} \sum_{j=1}^{2m} a_{ij} dx_i dx_j.$$

It has been known, since Poincaré's *Méthodes Nouvelles de la Mécanique Céleste*, vol. III, No. 247, that from an absolute invariant of order 2 can be deduced a set of others, of orders 4, 6, 8, $\cdots$, successively to the last of order $2m$. In the case $m=2$ by one operation we have the volume invariant

$$\int |a_{ij}|^{1/2} dx_1 dx_2 dx_3 dx_4,$$

where $|a_{ij}|$ is the skew-symmetric determinant of the $a_{ij}$. It is easy to verify this result by showing that the rational expression $|a_{ij}|^{1/2}$ is the "last multiplier" of the system. This verification suggests that the preceding result may be valid whatever may be the number $2m$ of the independent variables occurring in the system.

That is what we shall now demonstrate. We suppose, for example, that, in $a_{ij}$, $i$ indicates the rank of the row, $j$ that of the column, and we call $A_{ij}$ the minor corresponding to $a_{ij}$. The resolution of the system gives $2m$ equations:

$$\frac{dx_i}{dt} = \frac{A_{1i}\dfrac{\partial Z}{\partial x_1} + A_{2i}\dfrac{\partial Z}{\partial x_2} + \cdots + A_{ni}\dfrac{\partial Z}{\partial x_n}}{|a_{ij}|} = P_i.$$

We want to show that $|a_{ij}|^{1/2}$ is the last multiplier, so that

$$\sum_{i=1}^{2m} \frac{\partial(|a_{ij}|^{1/2} P_i)}{\partial x_i} = 0. \tag{1}$$

We shall remark, first, that in the preceding sum all the terms containing second derivatives of $Z$ vanish, because the $A_{ij}$ themselves form a skew-symmetric square array. Let us write

$$f_{ij} = \frac{A_{ij}}{|a_{ij}|^{1/2}}$$

and equate to 0 each of the coefficients of the terms involving

$$\frac{\partial Z}{\partial x_1}, \frac{\partial Z}{\partial x_2}, \cdots, \frac{\partial Z}{\partial x_n}$$

in the relation (1). Thus, we have the $2m$ relations

$$\sum_{j=1}^{2m} \frac{\partial f_{ij}}{\partial x_j} = 0.$$

If one of these relations is satisfied, on account of the symmetry of the indices, the same will be true for all the others. Then we shall limit ourselves to a verification of the first one of these relations.

Each term in this equation involves one second derivative of one of the $X_i$. This derivative in any arbitrary term can be found by taking for example in $f_{1s}$ $(s \neq 1)$ one factor $a_{uv}$ and in this factor $\partial X_u/\partial x_v$ which is to be derived with respect to $x_s$. So, an arbitrary term can be written

$$\frac{\partial^2 X_u}{\partial x_v \partial x_s} M,$$

the indices $u$, $v$, $s$ taking all the values between 1 and $2m$, but under the following condition: the four numbers 1, $s$, $u$, $v$ must all be distinct. The way in which this condition can be brought about is to be explained as follows: first, $s \neq 1$ because $f_{11} = 0$; then $f_{1s}$ is the minor corresponding to $a_{1s}$ in the "Pfaffian" $|a_{ij}|^{1/2}$ and in a "Pfaffian" each index can be written only once, so $u$, $v$ are entirely distinct from 1 and $s$; moreover, $u \neq v$ because $a_{uu} = 0$. We shall now show that there exists one term which is just the negative of the one we have written above; thus they will vanish together, and, since the first term was arbitrary, we can conclude that all the terms vanish.

The first equation with which we are dealing involves the expression $\partial f_{1v}/\partial x_v$ because $v \neq 1$. But $f_{1v}$ surely contains terms in which there is to be found the factor $a_{us}$ because all the numbers 1, $s$, $u$, $v$ are distinct. Hence, after differentiation we have a term such as

$$\frac{\partial^2 X_u}{\partial x_s \partial x_v} M'.$$

It remains to show that $M' = -M$. For that purpose we first remark that in the expansion of the whole determinant $|a_{ij}|$, $M|a_{ij}|^{1/2}$ can be considered as the coefficient of the product $a_{1s}a_{uv}$ and $M'|a_{ij}|^{1/2}$ as the coefficient of the product $a_{1v}a_{us}$. Now the truth of the assertion that one of these coefficients ($M|a_{ij}|^{1/2}$ and $M'|a_{ij}|^{1/2}$) is the negative of the other appears at once when we take the expansion of the determinant $|a_{ij}|$ according to Laplace's theorem written in the following form:

$$|a_{ij}| = \sum(-1)^{\nu} \begin{vmatrix} a_{1p} & a_{1q} \\ a_{up} & a_{uq} \end{vmatrix} \begin{vmatrix} a_{2,r} & a_{2,s} & \cdots & a_{2,t} \\ \cdots & \cdots & \cdots & \cdots \\ a_{u-1,r} & a_{u-1,s} & \cdots & a_{u-1,t} \\ a_{u+1,r} & a_{u+1,s} & \cdots & a_{u+1,t} \\ \cdots & \cdots & \cdots & \cdots \\ a_{2m,r} & a_{2m,s} & \cdots & a_{2m,t} \end{vmatrix}$$

$$(p, q, r, s, \cdots, t = 1, \cdots, 2m),$$

where we consider the minors of the second order composed of elements of the rows of rank 1 and $u$.

So the last of Poincaré's successive invariants can be written

$$\int |a_{ij}|^{1/2} dx_1 dx_2 \cdots dx_{2m}$$

and this form, interesting because of its simplicity, has been directly obtained by a simple verification.

By using exterior calculus we can present this result in a slightly different form. Referring to Cartan (*Leçons sur les Invariants Intégraux*, p. 78) we know that if $\int\Omega'$ is an absolute integral invariant ($\Omega' = \sum a_{ij}[dx_i dx_j]$), $\int(\Omega')^m$ is the absolute integral invariant of order $2m$. Hence what we have established above is equivalent to the formula

$$(\Omega')^m = K\,|a_{ij}|^{1/2} dx_1 \cdot dx_2 \cdots dx_{2m}. \tag{III}$$

This formula can be obtained also by merely using the rules of the exterior calculus. But whichever way one proceeds, one has to deal with the expressions called "Pfaffians"* in the theory of skew-symmetrical determinants, the definition of which can be stated as follows: $\sum(-1)^q a_{i_1 i_2} a_{i_3 i_4} \cdots a_{i_{2m-1} i_{2m}}$, the sum being extended to all the terms that are different, $q$ being the permutation number of $i_1, i_2, \cdots, i_{2m}$.

## II

By the use of Hamiltonian systems as a first step it is easy to see that the change of variables

$$x_1 = f_1\left(q, \frac{\partial L}{\partial q'}\right),$$
$$\cdots\cdots\cdots\cdots$$
$$x_{2m} = f_{2m}\left(q, \frac{\partial L}{\partial q'}\right)$$

* See Cayley, *Recherches ultérieures sur les déterminants gauches*, Crelle's Journal, vol. 55, pp. 299–313, and, for example, T. Muir, *The Theory of Determinants*, vol. II, London, 1911.

Also von Weber, *Vorlesungen über das Pfaff'sche Problem*, Leipzig, Teubner, 1900, where the Pfaffians are mentioned as "Pfaffsche Aggregat."

transforms any Lagrangian system, given by a function $L(q, q')$ in the variables $q$, $q'$, into a Pfaffian system in the variables $x_1, x_2, \cdots, x_{2m}$. Let us fix our attention for a while on the most general transformations by which we can pass from Lagrangian to Pfaffian systems. Suppose now that by means of the transformation

$$(2) \qquad q_i = F_i(x_1, x_2, \cdots, x_{2m}); \; q_i' = G_i(x_1, x_2, \cdots, x_{2m}) \qquad (i = 1, 2, \cdots, m)$$

the following equality

$$(3) \qquad L(F, G) = \sum_{j=1}^{2m} X_j(x_1, x_2, \cdots, x_{2m}) x_j' + Z(x_1, x_2, \cdots, x_{2m})$$

holds as long as the variables $x_j$ satisfy the system

$$(4) \qquad x_j' = \frac{\sum_{k=1}^{2m} A_{kj} \dfrac{\partial Z}{\partial x_k}}{\Delta} \qquad (j = 1, 2, \cdots, 2m).$$

In these relations (4), $x_j$, $Z$ are $2m+1$ functions of $x_1, x_2, \cdots, x_{2m}$; $\Delta$ is $|a_{ij}|$, the determinant of the

$$a_{ij} = \frac{\partial X_i}{\partial x_j} - \frac{\partial X_j}{\partial x_i},$$

and $A_{kj}$ is the minor corresponding to the element $a_{kj}$ in that determinant. Furthermore, we know that (4) gives the extremals ($\delta \int L = 0$) if we replace $L$ by the second member of the equation (3); so this system (4) will take the place of the Lagrangian equations after the transformation (2). The equation (3) must be satisfied only when we have (4), and for that the necessary and sufficient condition is

$$(5) \qquad L(F, G) - \frac{1}{\Delta} \sum_{j=1}^{2m} \left[ X_j \sum_{k=1}^{2m} A_{kj} \frac{\partial Z}{\partial x_k} \right] = Z.$$

Moreover, let us remark that the relations $q_i' = \partial q_i / \partial t$ give, in the new variables,

$$G_i = \sum_{j=1}^{2m} \frac{\partial F_i}{\partial x_j} x_j',$$

which are to be satisfied at the same time as (4); hence, we have in addition to (5) the following $m$ conditions:*

$$(6) \qquad \sum_{j=1}^{2m} \left[ \frac{\partial F_i}{\partial x_j} \sum_{k=1}^{2m} A_{kj} \frac{\partial Z}{\partial x_k} \right] = G_i \Delta.$$

* In (5) and (6) the fractions $A_{ij}/\Delta$ can be replaced by $f_{ji}/\Delta$ which are also rational expressions. The $f_{ij}$ are the minors of the Pfaffian $|\Delta|^{1/2}$ according to the definitions we have already recalled at the end of § I.

The $m+1$ relations (5) and (6) constitute a system which we call $(S)$ relating $F_i$, $G_i$, $X_i$, $Z$, all functions of $2m$ independent variables $x$ ($L$ being always a given function).

This system $(S)$ in which $F_i$, $G_i$ are considered as given functions and $X_i$, $Z$ as the unknown functions is in general completely integrable;* therefore, we conclude that by an appropriate choice of the $X_i$ and $Z$ a Lagrangian system can in general be reduced to a Pfaffian system by use of a transformation (2) in which $F_i$ and $G_i$ are all arbitrary. In a similar way we can assume the $X_i$, $Z$ given, and then we have a solution in $F_i$, $G_i$. *Hence, it follows that we can reduce any Lagrangian system to the Pfaffian system corresponding to a form $\omega$ arbitrarily chosen.*†

In particular in the case

$$\omega = y_{m+1}dy_1 + y_{m+2}dy_2 + \cdots + y_{2m}dy_m - y_1dy_{m+1} - y_2dy_{m+2} - \cdots - y_mdy_{2m} - 2y_{m+1}dt$$

we can, by the preceding general process, write the given system

$$\frac{dy_1}{dt} = 1, \quad \frac{dy_2}{dt} = 0, \quad \cdots, \quad \frac{dy_{2m}}{dt} = 0.$$

This is the "normal form" considered by Birkhoff,‡ who has shown that any Lagrangian system can be reduced to it in an actual sense by solving equations deduced from a variational principle. The method we give here is purely formal, but on the other hand it leads directly to the normal form, and thus the resolution with respect to $dy_1/dt$, $dy_2/dt$, $\cdots$, $dy_{2m}/dt$ of the differential equations of the extremals is rendered unnecessary.

From the most general point of view, we can say that if we consider as known a certain set of functions among the $F_i$, $G_i$, $X_i$, $Z$, the other functions can be given by the system $(S)$ under the condition that this system admit a solution (with respect to the unknown functions considered). As an example, it is easy to see that we obtain, by use of the system $(S)$, the reduction to the Hamiltonian form when we suppose given the $F_i$ and the $X_i$ in the following way:

---

* Indeed the system $(S)$ can, *in general*, be written in such a form that it will appear, according to Riquier's theory (*Les Systèmes d'Equations aux Dérivées Partielles*, Paris, Gauthier-Villars, 1910) to be "orthonome" and "passif," consequently *completely integrable*, i.e., admitting an existence theorem. When we say *"in general"* we may leave out of consideration some exceptional points. It is easy to see that these points are extremely uncommon.

† $\omega = \sum X_i dx_i + Zdt$. See first section.

‡ *Dynamical Systems*, loc. cit., p. 56.

$$F_1 = x_1, \quad F_2 = x_2, \quad \cdots, \quad F_m = x_m,$$
$$X_1 = x_{m+1}, \quad X_2 = x_{m+2}, \quad \cdots, \quad X_m = x_{2m},$$
$$X_{m+1} = 0, \quad X_{m+2} = 0, \quad \cdots, \quad X_{2m} = 0,$$

the $G_i$ and $Z$ being the only unknown functions.

## III

In this section we shall consider the Pfaffian systems formerly defined, from a different point of view, in order to show the use one can make of these systems in dynamical problems. First, for that purpose, we have to recall the fundamental results given for the first time by Birkhoff when he defined and introduced the consideration of those systems. In his book already cited* he demonstrates that in a *formal* sense "the Pfaffian equilibrium problem" can in general be reduced to Hamiltonian form, with the following consequence: "the normal form in the Hamiltonian case serves also in the Pfaffian case." An analogous statement can be made for generalized Pfaffian systems with respect to the "generalized Pfaffian equilibrium problem." According to these results there is to be found in the two following chapters of the same book a study of stability and existence of periodic motions based on the use of Pfaffian as well as Hamiltonian systems. Furthermore let us recall the following remark which shows the significance of the form of the dynamical equations for stability: the Pfaffian systems possess the same property as the Hamiltonian "of automatically fulfilling all of the conditions for complete stability, once the obvious conditions for first order stability are satisfied."

It is not without interest to obtain, from the considerations developed in the first section, the least order to which the system can be reduced in some classical problems of dynamics. The results are of course well known, but we shall give the following demonstration on account of its simplicity. In the problem of three bodies one starts with a Hamiltonian system of the eighteenth order for which one knows the energy integral $H=C$ and nine other integrals (all independent of each other). We can consider the system as Pfaffian, and according to a statement made above,† by the use of the nine integrals we can reduce its order by ten. After this reduction it will be of the same Pfaffian form and still admit the energy integral $H=C$. Then it will be of order 8, still reducible by two units more, by the use of $H=C$ and elimination of $t$.‡ This result can be generalized for the $n$-body problem for which

* Pp. 89–94.

† First section, p. 821.

‡ First section, p. 821.

the order of the system can be reduced to $6n-12$. In the case of the "three-body problem in a plane" by an analogous treatment the final order of the system can be made equal to 4.

In his memoir *Stability and the equations of dynamics* G. D. Birkhoff gives a special process which reduces a prepared linear differential form $R_1d\xi_1+R_2d\xi_2+\cdots+S_1d\eta_1+S_2d\eta_2+\cdots+Qdt$ corresponding to a Pfaffian system (in the vicinity of an equilibrium point for this system) to the sum of a canonical form $\omega$ and of a perfect differential $dE$, by the use of a succession of changes of variables such as

$$(C) \qquad \xi_i = \bar{\xi}_i + F_i, \qquad \eta_i = \bar{\eta}_i + G_i,$$

$F_i$, $G_i$ beginning with terms of at least the second degree.* But the reduction was only considered there in a "formal sense," so the question remained open whether or not it could be accomplished by convergent series. In the rest of this section we shall obtain, for the case of two degrees of freedom, an *actual* transformation, i.e., a change of variables of the form $(C)$ defining the considered reduction by means of analytic functions $F_i$, $G_i$. With Birkhoff's notations, $R_1$, $R_2$, $S_1$, $S_2$, $Q$ are functions of $\xi_1$, $\xi_2$, $\eta_1$, $\eta_2$; $R_1$, $R_2$, $Q$ begin with terms of the second order at least, whereas in $S_1$, $S_2$ the lower terms are respectively $\xi_1$, $\xi_2$. By the change of variables

$$(7) \qquad \begin{aligned} \xi_i &= \bar{\xi}_i, \\ \eta_i &= \bar{\eta}_i + G_i(\bar{\xi}_1, \bar{\xi}_2, \bar{\eta}_1, \bar{\eta}_2) \end{aligned}$$

($G_i$ beginning with terms of at least the second degree) the preceding form becomes

$$\begin{aligned} \bar{\omega} = {} & d\bar{\xi}_1\left[\bar{S}_1\frac{\partial G_1}{\partial \bar{\xi}_1} + \bar{S}_2\frac{\partial G_2}{\partial \bar{\xi}_1} + \bar{R}_1\right] \\ & + d\bar{\xi}_2\left[\bar{S}_1\frac{\partial G_1}{\partial \bar{\xi}_2} + \bar{S}_2\frac{\partial G_2}{\partial \bar{\xi}_2} + \bar{R}_2\right] \\ & + d\bar{\eta}_1\left[\bar{S}_1\frac{\partial G_1}{\partial \bar{\eta}_1} + \bar{S}_2\frac{\partial G_2}{\partial \bar{\eta}_1} + \bar{S}_1\right] \\ & + d\bar{\eta}_2\left[\bar{S}_1\frac{\partial G_1}{\partial \bar{\eta}_2} + \bar{S}_2\frac{\partial G_2}{\partial \bar{\eta}_2} + \bar{S}_2\right] + \bar{Q}dt. \end{aligned}$$

* If we consider only the reduction of a Pfaffian system to a Hamiltonian system it is sufficient to reduce $\omega+dE$ to a canonical form $\bar{\omega}$.

Properly choosing $dE=\sum c_i dx_i$ we can be sure that all the coefficients of $\omega^{(2m-2)}$, i.e. of $(\omega')^{m-1}\omega$, vanish only if the same is true of the coefficients of $(\omega')^{m-1}$ and therefore of the coefficients of $(\omega')^m$. The fact that all these coefficients do not vanish in our problem, even in the vicinity of an equilibrium point, is a result of the condition $|a_{ij}|\neq 0$ and of the formula (III). Then the possibility of an actual reduction follows as a consequence of Goursat's theorem stating the existence of *conjugate groups of functions*: *Le Problème de Pfaff*, Paris, Gauthier-Villars, p. 186.

Suppose now that the following system, in which $E$ is any function whatever of $\xi_1, \xi_2, \eta_1, \eta_2$ (we now omit the bars)

$$(8)\qquad \begin{aligned} \frac{\partial E}{\partial \xi_1} &= S_1\frac{\partial G_1}{\partial \xi_1} + S_2\frac{\partial G_2}{\partial \xi_1} + R_1, \\ \frac{\partial E}{\partial \xi_2} &= S_1\frac{\partial G_1}{\partial \xi_2} + S_2\frac{\partial G_2}{\partial \xi_2} + R_2 \end{aligned}$$

admits a solution $E, G_1, G_2$. Then by subtracting the exact differential

$$\frac{\partial E}{\partial \xi_1}d\xi_1 + \frac{\partial E}{\partial \xi_2}d\xi_2 + \frac{\partial E}{\partial \eta_1}d\eta_1 + \frac{\partial E}{\partial \eta_2}d\eta_2$$

from the preceding sum $\tilde{\omega}$ we see it is sufficient to take for $G_1, G_2$ a solution of (8) in order to make vanish the coefficients of the terms in $d\xi_1$ and $d\xi_2$ in the above expression of $\tilde{\omega}$. It remains now only to establish the existence of a solution for the system (8).

Let us change our notation; with

$$\begin{aligned} E &= u_1, \quad G_1 = u_2, \quad G_2 = u_3, \\ \xi_1 &= x_1, \quad \xi_2 = x_2, \quad \eta_1 = x_3, \quad \eta_2 = x_4, \end{aligned}$$

the system (8) can be written

$$\frac{\partial u_1}{\partial x_1} = S_1\frac{\partial u_2}{\partial x_1} + S_2\frac{\partial u_3}{\partial x_1} + R_1, \quad \frac{\partial u_1}{\partial x_2} = S_1\frac{\partial u_2}{\partial x_2} + S_2\frac{\partial u_3}{\partial x_2} + R_2,$$

$S_1, S_2, R_1, R_2$ being now expressed by means of the variables $x_1, x_2, x_3, x_4$ and the two unknown functions $u_2, u_3$.

Writing the condition of integrability for this system and introducing two new functions by

$$\frac{\partial u_2}{\partial x_1} = u_4, \quad \frac{\partial u_2}{\partial x_2} = u_5,$$

we have

$$\frac{\partial u_1}{\partial x_1} = S_1u_4 + S_2\frac{\partial u_3}{\partial x_1} + R_1, \quad \frac{\partial u_1}{\partial x_2} = S_1u_5 + S_2\frac{\partial u_3}{\partial x_2} + R_2, \quad \frac{\partial u_2}{\partial x_1} = u_4, \quad \frac{\partial u_2}{\partial x_2} = u_5,$$

$$(9)\qquad \frac{\partial u_3}{\partial x_1} = \frac{\left[u_4\left(\frac{\partial S_2}{\partial u_2} - \frac{\partial S_1}{\partial u_3}\right) + \frac{\partial S_2}{\partial x_1} - \frac{\partial R_1}{\partial u_3}\right]\frac{\partial u_3}{\partial x_2} - u_4\left(\frac{\partial S_1}{\partial x_2} - \frac{\partial R_2}{\partial u_2}\right) + u_5\left(\frac{\partial S_1}{\partial x_1} - \frac{\partial R_1}{\partial u_2}\right) - \frac{\partial R_1}{\partial x_2} + \frac{\partial R_2}{\partial x_1}}{\frac{\partial S_2}{\partial x_2} - \frac{\partial R_2}{\partial u_3} + u_5\left(\frac{\partial S_2}{\partial u_2} - \frac{\partial S_1}{\partial u_3}\right)},$$

$$\frac{\partial u_5}{\partial x_1} = \frac{\partial u_4}{\partial x_2}.$$

For the purpose of studying this linear system of simultaneous partial differential equations of the first order we can refer to C. Bourlet's thesis* *Sur les équations aux dérivées partielles simultanées* and especially to the second part of it devoted to the case of linear systems; it is clear then that (9) is of the form

$$\frac{\partial u_i}{\partial x_k} = a_{00}^{ik} + \sum_{j>i} a_{jk}^{ik} \frac{\partial u_j}{\partial x_k} + \sum_{h>k} a_{ih}^{ik} \frac{\partial u_h}{\partial x_h},$$

where the $a$ are functions of the $x$ and $u$. So this linear system belongs to the class denoted there as "canonique" and moreover it is *completely integrable*, i.e. the two expressions of every "doubly principal derivative" can be shown to be identical by the use of the system itself and of the equations that can be deduced from it by differentiation. If we remark that the coefficients $a$ are in our case analytic in the vicinity of $x_1=x_2=x_3=x_4=0$, $u_1=u_2=u_3=u_4=u_5=0$, Theorem VIII of Bourlet can be applied to our system. Hence we conclude that (9) admits a solution analytic near $x_1=x_2=x_3=x_4=0$, and for which $u_1=u_2=u_3=u_4=u_5=0$. This solution is to be obtained by taking all the "parametric derivatives" equal to 0. Furthermore in the case where all the parametric derivatives are equal to 0, we observe that for the solution all the first derivatives vanish. Hence the $G_i$ do not involve terms of the first order, and after the change of variables (7) determined by the solution under consideration the function $\overline{Q}$ in $\bar{\omega}$ still begins with terms of the second order at least and we are still in the vicinity of an equilibrium point.

We achieve the reduction to the Hamiltonian form by

$$\begin{aligned} \eta_i &= \bar{\eta}_i, \\ S_1\frac{\partial G_1}{\partial \eta_1} + S_2\frac{\partial G_2}{\partial \eta_1} + S_1 - \frac{\partial E}{\partial \eta_1} &= \bar{\xi}_1, \\ S_1\frac{\partial G_1}{\partial \eta_2} + S_2\frac{\partial G_2}{\partial \eta_2} + S_2 - \frac{\partial E}{\partial \eta_2} &= \bar{\xi}_2. \end{aligned} \tag{10}$$

The two last equations can be solved so as to give $\xi_1$, $\xi_2$ in the vicinity of the origin, the Jacobian $D(\bar{\xi}_1, \bar{\xi}_2)/D(\xi_1, \xi_2)$ being there equal to 1, since the $\partial^2 E/\partial\eta_i\partial\xi_k$ vanish, as is easy to verify by differentiating (8).

If we pay special attention to the terms of lower degree, the resolution of the equations (10) gives

$$\begin{aligned} \xi_1 &= \bar{\xi}_1 + \cdots, \\ \xi_2 &= \bar{\xi}_2 + \cdots, \end{aligned}$$

* Also Annales de l'Ecole Normale Supérieure. (3), vol. 8 (1891), supplément, p. 1.

and that enables us to state that we remain in the vicinity of an equilibrium point. So by the successive use of the two changes of variables (7) and (10) in which $G_1$, $G_2$ are a solution of the system (8), analytic in the vicinity of $x_i=0$, $u_i=0$ we can reduce the expression $\omega$ to $\xi_1 d\eta_1+\xi_2 d\eta_2+Qdt$ in which $Q$ begins with terms of the second order at least. In conclusion we have accomplished in a *formal sense*, i.e. *by use of convergent series*, the reduction which is proposed in Birkhoff's memoir. The prepared linear form from which we start has been transformed by a change of variables $(C)$ into the sum of a canonical $\omega$ and of a perfect differential $dE$.†

The way we have proceeded suggests the following remarks:

(1) $dE$ itself vanishes in the vicinity of the point under consideration;

(2) the unknown functions $F_i$, $G_i$ which determine the change of variables are given by a system of partial differential equations, which system is linear, and completely integrable, and which we have written out entirely;

(3) in the system (9), $\eta_1$, $\eta_2$ are parametric variables for the function $G_1$, so every term of the solution $G_1$ must involve either $\xi_1$, $\xi_2$ or both, and

$$G_1 = \xi_1 G_1^*(\xi_1\xi_2\eta_1\eta_2) + \xi_2 G_1^{**}(\xi_1\xi_2\eta_1\eta_2).$$

In the same way, $\xi_2$, $\eta_1$, $\eta_2$ being parametric variables with respect to $G_2$ for the solution of (9), we can write

$$G_2 = \xi_1 G_2^*(\xi_1\xi_2\eta_1\eta_2).$$

It follows that the solution whose existence we have established both in a formal and in an actual sense, is different from the formal solution that can be obtained by a succession of changes of variables according to the general method of Birkhoff.

† This conclusion, the main result of §3, has been extended, during the printing of this paper, to the general case, i.e., for any number of degrees of freedom. Cf. Paris Comptes Rendus, vol. 190, p. 358. The method used for this extension is almost the same as the one explained in detail above.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# ON CERTAIN IDENTITIES IN THETA FUNCTIONS*

BY

CLAIBORNE G. LATIMER

1. The purpose of this paper is to obtain certain identities in Jacobi's theta functions by a new method. This method may be outlined as follows. We consider two sets of elements $S$ and $S_1$, conceptually distinct, such that the elements of $S$ are in one-to-one correspondence with the elements of $S_1$. In certain cases these elements are algebraic numbers and in other cases they are quaternions; in every case, corresponding elements have the same norms. We construct two functions $\Phi(x, y, \cdots)$ and $\Psi(x_1, y_1, \cdots)$, the terms of which are in one-to-one correspondence with the elements of $S$ and $S_1$ respectively. Hence there is a one-to-one correspondence between the terms of $\Phi$ and those of $\Psi$. We then show that if $x_1, y_1, \cdots$ are properly defined linear functions of $x, y, \cdots$, the corresponding terms are equal and therefore $\Phi(x, y, \cdots) = \Psi(x_1, y_1, \cdots)$.

We shall prove Jacobi's well known fundamental identity,† i.e.

$$\Phi(x, y, z, w) \equiv \vartheta_3(x)\vartheta_3(y)\vartheta_3(z)\vartheta_3(w) + \vartheta_2(x)\vartheta_2(y)\vartheta_2(z)\vartheta_2(w) = \Phi(x_1, y_1, z_1, w_1),$$

where $x_1, y_1, z_1, w_1$ are certain linear functions of $x, y, z, w$. It will be shown that the terms of $\Phi$ are in one-to-one correspondence with the set of all (Hurwitz) integral quaternions.‡ The above identity is then proved by employing certain well known properties of these quaternions. A generalization of this identity is obtained in a similar manner by employing sets of integral elements in generalized quaternion algebras. These sets were determined by the writer in a former paper,§ employing a definition of integral element due to Dickson.¶

The other identities are obtained by employing sets of elements not integral in general. We also obtain Schroeter's identity by the same method.

---

* Presented to the Society, March 30, 1929, and December 27, 1929; received by the editors in October, 1929.

† Cf. Tannery and Molk, *Eléments de la Théorie des Fonctions Elliptiques*, vol. 2, p. 162. Kronecker's notation for the theta functions will be used throughout this paper.

‡ Hurwitz, *Vorlesungen über die Zahlentheorie der Quaternionen.*

§ American Journal of Mathematics, vol. 48 (1926), pp. 57–66. We shall hereafter refer to this paper as J.

¶ *Algebras and their Arithmetics*, pp. 141, 2, 6. In J, Dickson's property $U'$ was used instead of $U$. In the algebra of rational quaternions, the integral elements under the definition used in J are the same as the Hurwitz integral elements.

The use of algebraic numbers and quaternions, and in particular their arithmetics, to obtain identities in theta functions is believed to be new and not without interest.

2. **Proof of Jacobi's fundamental identity.** Part of the argument used to establish this identity may be repeated verbatim later on to prove a generalization.

We shall recall certain well known definitions and properties of real quaternions. If $X=x+yi+zj+wk$ is a real quaternion, where $i, j, k$ are the usual quaternion units, the conjugate of $X$ is $X'=2x-X$ and the norm of $X$ is $N(X)=XX'=x^2+y^2+z^2+w^2$. Then $N(XY)=N(X)N(Y)$ and $(XY)'=Y'X'$.

In the algebra of rational quaternions, the (Hurwitz) integral quaternions are those in the form

$$\tfrac{1}{2}(\xi+\eta i+\zeta j+\lambda k)$$

where $\xi, \eta, \zeta, \lambda$ are integers which are all even or all odd. The set $S$ of all such quaternions is closed under multiplication.

By definition*

$$\vartheta_2(x, q) = \sum q^{(2m+1)^2/4} e^{(2m+1)xi},$$
$$\vartheta_3(x, q) = \sum q^{m^2} e^{2mxi},$$

where $q$ is a complex number, $|q|<1$. Since these functions are absolutely convergent we have, dropping the parameter $q$,

$$\vartheta_j(x)\vartheta_j(y)\vartheta_j(z)\vartheta_j(w) = \sum q^{(\xi^2+\eta^2+\zeta^2+\lambda^2)/4} e^{(\xi x+\eta y+\zeta z+\lambda w)i},$$

where for $j=2$ [$j=3$] the summation extends over all systems of four odd [even] integers $\xi, \eta, \zeta, \lambda$. Assume for the present that $x, y, z, w$ are real and let

$$\Omega = \tfrac{1}{2}(\xi+\eta i+\zeta j+\lambda k), \qquad X = x+yi+zj+wk.$$

Let the real part of the quaternion $Y$ be designated by $R(Y)$. Then

$$\Phi(x, y, z, w) = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where $\Phi$ is as defined in §1 and the summation extends, without duplication or omission, over all elements of $S$. The terms of $\Phi$ are thus in one-to-one correspondence with the elements of $S$.

Let $U$ be an element of $S$ of norm unity and let $x_1, y_1, z_1, w_1$ be defined in terms of $x, y, z, w$ by

$$X_1 = XU,$$

* It will be understood throughout this paper that the limits of summation are $-\infty$ and $+\infty$ unless otherwise stated.

where

$$X_1 = x_1 + y_1 i + z_1 j + w_1 k.$$

Let $\Omega$ be an arbitrarily chosen element of $S$ and let $\tau$ be the corresponding term of $\Phi(x, y, z, w)$. Since $S$ is closed under multiplication,

$$\Omega U = \Omega_1$$

belongs to $S$. Let $\tau_1$ be the term of $\Phi(x_1, y_1, z_1, w_1)$ which corresponds to $\Omega_1$. We have

$$N(\Omega_1) = N(\Omega)N(U) = N(\Omega),$$
$$\Omega_1 X_1' = (\Omega U)(XU)' = (\Omega U)(U'X') = \Omega X'.$$

Hence

$$\tau = q^{N(\Omega)} e^{2\pi i R(\Omega X')} = q^{N(\Omega_1)} e^{2\pi i R(\Omega_1 X_1')} = \tau_1.$$

Since $\tau = \tau_1$ is a formal identity in $x, y, z, w$, we may discard our assumption that they are real. $U$ belongs to $S$ and therefore the same is true of $U'$. Hence by means of $\Omega U = \Omega_1$, we have a one-to-one correspondence between the terms of $\Phi(x, y, z, w)$ and $\Phi(x_1, y_1, z_1, w_1)$. And, as we have just seen, corresponding terms are equal. Therefore

$$\Phi(x, y, z, w) = \Phi(x_1, y_1, z_1, w_1). \tag{1}$$

If we take as $U$ one of the elements

$$\tfrac{1}{2}(\pm 1 \pm i \pm j \pm k)$$

and change the signs of certain of the variables, (1) becomes Jacobi's identity. The only other elements in $S$ of norm unity are $\pm 1, \pm i, \pm j, \pm k$. If we take one of these as $U$, (1) is trivial.

3. **The sets of integral elements in generalized quaternion algebras.** Consider the rational algebra $A$, with the basal numbers

$$1, \quad I = (\beta\gamma)^{1/2} i, \quad J = (\gamma\alpha)^{1/2} j, \quad K = (\alpha\beta)^{1/2} k,$$

where $\alpha, \beta, \gamma$ are positive odd integers such that $\alpha\beta\gamma$ contains no square factor $>1$. $A$ is a subalgebra of the algebra of real quaternions and hence has the properties mentioned in §2.*

Two of $\alpha, \beta, \gamma$ are congruent, modulo 4, and therefore, after applying a cyclic substitution if necessary, we may assume that $\alpha \equiv \beta \pmod 4$. An integral element of $A$ is defined as an element which belongs to a set $S$ which

* In J we considered an algebra $A_1$, with a multiplication table which involved certain parameters $\alpha$ and $\beta$. Let $\alpha = \alpha_1\delta, \beta = \beta_1\delta$, where $\delta$ is the g.c.d. of $\alpha$ and $\beta$, and assume that $\alpha < 0$ and $\beta > 0$. If in the multiplication table of $A$ we replace $\alpha_1, \beta_1, \delta, K$ by $-\beta, \alpha, \gamma, \gamma K$ respectively, it will be seen that $A_1$ is equivalent to $A$. We shall quote the results of J in terms of the new notation.

has certain properties $R$, $C$, $U$, $M$.* By property $C$, if $X$, $Y$ are elements of $S$ then $XY$ also belongs to $S$. We shall not employ explicitly the properties $R$, $U$, $M$.

Let $\alpha = A\alpha'$, $\beta = B\beta'$, $\gamma = C\gamma'$ where $A$, $B$, $C$ are defined as follows. $A$ is the product of all the prime factors $A_i$ of $\alpha$ such that $-\beta\gamma$ is a quadratic non-residue of every $A_i$; or $A = 1$ if $\alpha$ contains no such factor. $B$ and $C$ are similarly defined with a cyclic substitution on $A$, $B$, $C$ and on $\alpha$, $\beta$, $\gamma$. If $Q$ is the number of prime factors in $\alpha'\,\beta'\,\gamma'$, there are exactly $2^Q = t$ odd integers $H_i (i = 1, 2, \cdots, t)$, incongruent modulo $\alpha'\beta'\gamma'$ such that

$$(2)\quad \beta\gamma H_i^2 + 1 \equiv 0 \pmod{\alpha'},\ \gamma\alpha H_i^2 + 1 \equiv 0 \pmod{\beta'},\ \alpha\beta H_i^2 + 1 \equiv 0 \pmod{\gamma'}.$$

Let $H$ be one of the $H_i$ and let $S_i[S_i', S_i'']$ be the totality of elements in the form

$$(3)\quad \Omega = \frac{1}{2}\left(\xi + \frac{\eta}{\beta'\gamma'}\,I + \frac{\zeta}{\gamma'\alpha'}\,J + \frac{\lambda}{\alpha'\beta'}\,K\right),$$

where $\xi$, $\eta$, $\zeta$, $\lambda$ are rational integers such that

$$(4)\quad \begin{aligned} \zeta &\equiv B\gamma' H\lambda && \pmod{\alpha'},\\ \eta &\equiv A\gamma' H\lambda && \pmod{\beta'},\\ \lambda &\equiv A\beta' H\zeta && \pmod{\gamma'}, \end{aligned}$$

and such that the first [second, third] of the following conditions is satisfied:

$$\begin{aligned} &\xi \equiv \eta \equiv \zeta \equiv \lambda,\\ &\xi \equiv \eta, \quad \zeta \equiv \lambda,\\ &\xi \equiv \zeta, \quad \eta \equiv \lambda \end{aligned} \qquad \pmod 2.$$

We shall say that an element in the form (3) has the coördinates $\xi$, $\eta$, $\zeta$, $\lambda$. It will be noted that $S_i$ is the intersection of $S_i'$ and $S_i''$.

If $\alpha \equiv \gamma \pmod 4$, by the proof of Theorem 3, J, $S_i$ has the properties $R, C, U, M$. On the other hand, if $\alpha \equiv \gamma + 2 \pmod 4$, the sets $S_i'$ and $S_i''$ have these properties. For every $H_i$ we thus have either one or two sets of integral elements.

From (2) it follows that (4) is equivalent to

$$(5)\quad \begin{aligned} \lambda &\equiv -\beta' C H\zeta && \pmod{\alpha'},\\ \lambda &\equiv -\alpha' C H\eta && \pmod{\beta'},\\ \zeta &\equiv -\alpha' B H\eta && \pmod{\gamma'}. \end{aligned}$$

* J, p. 57.

If $\eta, \zeta, \lambda$ satisfy (4) and if $\zeta \equiv \lambda$ or $\eta \equiv \lambda \pmod 2$, by employing (5) it may be shown that they may be written in the corresponding forms below:

(a) $\zeta \equiv \lambda \pmod 2$:

$$\begin{aligned} \eta &= 2\beta'\gamma' m_1 + r, \\ \zeta &= 2\gamma'\alpha' m_2 - \alpha' B H r + \beta\gamma' H p, \\ \lambda &= 2\alpha'\beta' m_3 - \alpha' C H r + \beta' p, \\ 0 &\leqq p < 2\alpha', \quad 0 \leqq r < 2\beta'\gamma'; \end{aligned} \tag{6}$$

(b) $\eta \equiv \lambda \pmod 2$:

$$\begin{aligned} \eta &= 2\beta'\gamma' m_1 + A\beta' H r + \alpha\gamma' H p, \\ \zeta &= 2\gamma'\alpha' m_2 + r, \\ \lambda &= 2\alpha'\beta' m_3 - \beta' C H r + \alpha' p, \\ 0 &\leqq p < 2\beta', \quad 0 \leqq r < 2\alpha'\gamma', \end{aligned}$$

where $m_1, m_2, m_3, r, p$ are integers, $r$ and $p$ subject to the indicated inequalities. Conversely, if $\eta, \zeta, \lambda$ may be written in the forms (6a) [(6b)], they satisfy (4) and $\zeta \equiv \lambda$ [$\eta \equiv \lambda$] (mod 2).

4. **Construction of functions corresponding to the sets $S_i, S_i', S_i''$.** Let $q$ be a complex number $|q| < 1$; $a = A/\alpha'$, $b = B/\beta'$, $c = C/\gamma'$, and

$$\begin{aligned} F_1(x, r) &= \sum_{m_1} q^{bc\eta^2/4} e^{bc\eta x i}, \\ G_1(x; r, p) &= \sum_{m_2} q^{ca\zeta^2/4} e^{ca\zeta x i}, \\ H_1(x; r, p) &= \sum_{m_3} q^{ab\lambda^2/4} e^{ab\lambda x i}, \end{aligned}$$

where $r, p$ are integers and $\eta, \zeta, \lambda$ are given by (6a). Let $F_2(x; r, p)$, $G_2(x, r)$, $H_2(x; r, p)$ be defined in the same way except that we use the expressions (6b) for $\eta, \zeta, \lambda$. We may write

$$F_1(x, r) = q^{bcr^2/4} e^{bcrxi} \vartheta_3[BC(x - rk\hat{\imath}), q^{\beta\gamma}],$$

where $k = (\log q)/2$. The other five functions may be similarly expressed in terms of theta functions. Hence each of these six functions is absolutely convergent. Then

$$\begin{aligned} \Psi_0(s, t) &= \vartheta_3(x) F_1(y, 2s) G_1(z; 2s, 2t) H_1(w; 2s, 2t) \\ &= \sum q^{(\xi^2 + bc\eta^2 + ca\zeta^2 + ab\lambda^2)} e^{(\xi x + bc\eta y + ca\zeta z + ab\lambda w)i}, \end{aligned}$$

where the summation extends over all systems of four even integers $\xi, \eta, \zeta, \lambda$ such that $\eta, \zeta, \lambda$ may be written in the forms (6a) with $r, p$ replaced by $2s, 2t$ respectively. For every such system, there is a corresponding element of $S_i$,

$$\Omega = \frac{1}{2}\left(\xi + \frac{\eta}{\beta'\gamma'} I + \frac{\zeta}{\gamma'\alpha'} J + \frac{\lambda}{\alpha'\beta'} K\right).$$

Conversely, if $\Omega$ is an element of $S_i$ with even coördinates, $\xi, \eta, \zeta, \lambda$, there is a uniquely determined pair of integers $s, t$, $0 \leqq s < \beta'\gamma'$, $0 \leqq t < \alpha'$, such that $\Psi_0(s, t)$ contains exactly one term corresponding to $\Omega$. Let

$$\Psi_1(s, t) \equiv \vartheta_2(x) F_1(y, 2s + 1) G_1(z; 2s + 1, 2t) H_1(w; 2s + 1, 2t).$$

The above argument regarding $\Psi_0$ may be repeated for $\Psi_1$ except that $\xi, \eta, \zeta, \lambda$ are odd, $r$ is replaced by $2s+1$, and $\Omega$ has odd coördinates. Hence

$$\Phi_1(x, y, z, w) \equiv \sum_{s=0}^{\beta'\gamma'-1} \sum_{t=0}^{\alpha'-1} (\Psi_0(s, t) + \Psi_1(s, t))$$

$$= \sum_{\Omega} q^{N(\Omega)} e^{2R(\Omega X')i}, \tag{7}$$

where we assume for the present that $x, y, z, w$ are real,

$$X = x + \frac{y}{\beta'\gamma'} I + \frac{z}{\gamma'\alpha'} J + \frac{w}{\alpha'\beta'} K,$$

and the summation extends, without duplication or omission, over all elements of $S_i$. The terms of $\Phi_1$ are thus in one-to-one correspondence with the elements of $S_i$.

We shall next construct functions $\Phi_2$ and $\Phi_3$, the terms of which are in one-to-one correspondence with the elements of $S_i'$ and $S_i''$ respectively. As noted before, both $S_i'$ and $S_i''$ contain $S_i$. Accordingly, we shall obtain $\Phi_2$ and $\Phi_3$ by adding properly chosen functions to $\Phi_1$.

The $\eta, \zeta, \lambda$ coördinates of every element of $S_i'$ may be written in the forms (6a) and $\xi \equiv \eta$, $\zeta \equiv \lambda \pmod 2$. Furthermore, if such an element does not belong to $S_i$, $\eta \equiv \lambda + 1 \pmod 2$, and therefore in (6a) $r \equiv p + 1 \pmod 2$. Therefore if we set

$$\phi_{20} \equiv \vartheta_3(x) \sum_{s=0}^{\beta'\gamma'-1} \sum_{t=0}^{\alpha'-1} F_1(y, 2s) G_1(z; 2s, 2t + 1) H_1(w; 2s, 2t + 1),$$

$$\phi_{21} \equiv \vartheta_2(x) \sum_{s=0}^{\beta'\gamma'-1} \sum_{t=0}^{\alpha'-1} F_1(y, 2s+1) G_1(z; 2s+1, 2t+1) H_1(w; 2s+1, 2t+1),$$

it follows that $\phi_{20} + \phi_{21}$ is equal to the expression on the right of (7), where in this case the summation extends over all the elements of $S_i'$ which do not belong to $S_i$. Then

$$\Phi_2(x, y, z, w) \equiv \Phi_1 + \phi_{20} + \phi_{21} = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where the summation extends over all the elements of $S_i'$. If we let

$$\phi_{30} \equiv \vartheta_3(x) \sum_{s=0}^{\gamma'\alpha'-1} \sum_{t=0}^{\beta'-1} F_2(y;\, 2s,\, 2t+1) G_2(z,\, 2s) H_2(w;\, 2s,\, 2t+1),$$

$$\phi_{31} \equiv \vartheta_2(x) \sum_{s=0}^{\gamma'\alpha'-1} \sum_{t=0}^{\beta'-1} F_2(y;\, 2s+1,\, 2t+1) G_2(z,\, 2s+1) H_2(w;\, 2s+1,\, 2t+1),$$

it may be shown as for $\Phi_2$ that

$$\Phi_3(x, y, z, w) \equiv \Phi_1 + \phi_{30} + \phi_{31} = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where now the summation extends over all the elements of $S_i''$.

5. **Identities in functions which correspond to sets of integral elements.** For each of the sets $S_i, S_i', S_i''$ we have a corresponding function $\Phi_j (j = 1, 2, 3)$. As noted in §3, $S_i$ has the properties $R, C, U, M$ if $\alpha \equiv \gamma \pmod 4$, while $S'$ and $S''$ have these properties if $\alpha \equiv \gamma + 2 \pmod 4$. Let $S$ be one of the sets which has these properties and let $\Phi$ be the corresponding function. Let

$$X_1 = x_1 + \frac{y_1}{\beta'\gamma'} I + \frac{z_1}{\gamma'\alpha'} J + \frac{w_1}{\alpha'\beta'} K.$$

The argument of §2 beginning "Let $U$ be an element of $S \cdots$" and leading to (1) may now be repeated word for word, it being understood that we employ the $\Phi$, $X_1$ and $S$ of this paragraph and the $X$ and $\Omega$ of the preceding paragraph. We have then

THEOREM 1. *Let $S$ be one of the sets $S_i, S_i', S_i''$ which has the properties $R, C, U, M$, and let $\Phi$ be the corresponding $\Phi_j$. If $U$ is an element of $S$ of norm unity and if $x_1, y_1, z_1, w_1$ are defined in terms of $x, y, z, w$ by means of $XU = X_1$, then**

$$\Phi(x, y, z, w) = \Phi(x_1, y_1, z_1, w_1).$$

6. **Special cases of Theorem 1.** If $\alpha, \beta, \gamma$ are assigned numerical values, to write out explicitly the identity of the above theorem it is necessary to determine an element $U$ as there defined, i.e. a unit belonging to $S$. In some cases a unit yields a trivial identity, as was seen at the end of §2. Furthermore, the algebra may have no unit which yields a non-trivial identity. This is the case if $\alpha = \beta = 1, \gamma = 7$; the only units being $\pm 1, \pm K$.

By the proofs of Theorems 2 and 3 of J, every element with integral coördinates and integral norm belongs to a set of integral elements. The

* We have considered only those algebras in which $\alpha, \beta, \gamma$ are odd. The same method is being used in a Kentucky thesis, employing the algebras in which one of these parameters is even and the results of Professor Darkow on the arithmetics of such algebras. See Annals of Mathematics, vol. 28, pp. 263–70.

problem of determining those algebras which contain units yielding non-trivial identities is therefore equivalent to that of determining those values of $\alpha, \beta, \gamma$ such that the equation

$$\alpha'\beta'\gamma'\xi^2 + \alpha' BC\eta^2 + \beta' CA\zeta^2 + \gamma' AB\lambda^2 = 4\alpha'\beta'\gamma'$$

has a solution in integers, two of which are $\neq 0$. This, of course, is a very difficult problem. We shall give two classes of algebras in which such a unit may be determined and the corresponding identity written out explicitly.

Let $\beta=\gamma=1$ and let $\alpha$ be a product of distinct primes, each in the form $4n+1$. Then $A=B=C=\beta'=\gamma'=1$, $\alpha=\alpha'=u^2+v^2$, where $u, v$ are relatively prime. We may take

$$U = \frac{1}{2}\left(1 + I + \frac{u+v}{\alpha}J + \frac{u-v}{\alpha}K\right).$$

For example, if $\alpha=5$, $u=2$, $v=1$, $H=3$, it will be found that $\Phi$ may be written

$$\Phi(x, y, z, w) = \vartheta_3(x, q)\vartheta_3(y, q)\sum_{t=0}^{4} q^{2t^2} e^{2(3z+w)ti/5}\vartheta_3(z - 6tki, q^5)\vartheta_3(w - 2tki, q^5)$$

$$+ \vartheta_2(x, q)\vartheta_2(y, q)\sum_{t=0}^{4} q^{2t^2} e^{2(3z+w)ti/5}\vartheta_2(z - 6tki, q^5)\vartheta_2(w - 2tki, q^5).$$

By the theorem,

$$\Phi(x, y, z, w) = \Phi(x_1, y_1, z_1, w_1),$$

where we find

$$x_1 = \frac{1}{2}\left(x - y - \frac{3z+w}{5}\right), \quad z_1 = \frac{1}{2}(3x - y + z + w),$$

$$y_1 = \frac{1}{2}\left(x + y + \frac{z-3w}{5}\right), \quad w_1 = \frac{1}{2}(x + 3y - z + w).$$

Similar results may be obtained if $\alpha$ is a product of distinct primes, each in the form $3n+1$, one of $\beta, \gamma$ is unity and the other is 3, noting that the notation must be so chosen that $\alpha \equiv \beta \pmod 4$.

If the method of the preceding paragraphs be applied to the algebraic integers in the field defined by an imaginary cube root of unity and if we set

$$\Phi(x, y) \equiv \vartheta_3(x, q)\vartheta_3(y, q^3) + \vartheta_2(x, q)\vartheta_2(y, q^3),$$

we obtain the known identity

$$\Phi(x, y) = \Phi(x_1, y_1),$$

where

$$x_1 = \tfrac{1}{2}(x + y), \quad y_1 = \tfrac{1}{2}(3x - y).$$

7. **Proof of Schroeter's identity.** The sets of elements used hereafter will not be integral in general. Let $r$, $p$ be positive integers and let $s$, $t$ be integers such that $t$ is prime to

$$\Delta \equiv rs^2 + pt^2.$$

Let $\rho = r^{1/2}$, $\tau = ip^{1/2}$, and let $S$ be the totality of numbers in the form

$$\Omega = \Delta^{-1/2}(\xi + \eta\rho\tau),$$

where $\xi$, $\eta$ are integers such that

$$t\xi \equiv -rs\eta \pmod{\Delta}. \tag{8}$$

Let $S_1$ be the the totality of numbers in the form

$$\Omega_1 = \xi_1\rho + \eta_1\tau,$$

where $\xi_1$, $\eta_1$ are integers. If $U \equiv \Delta^{-1/2}(s\rho + t\tau)$ and

$$\Omega U = \Omega_1,$$

it may be shown that $\Omega_1$ belongs to $S_1$ if and only if $\Omega$ belongs to $S$. The numbers of $S$ are therefore in one-to-one correspondence with the numbers of $S_1$.

It may be shown that if $\mu$ is an integer and if

$$\Phi_\mu(x, y) \equiv q^{r\mu^2} e^{2r(sx - pty)\mu i} \vartheta_3(\Delta x - rs\mu ki, q^\Delta)\, \vartheta_3(pr\Delta y + prt\mu ki, q^{pr\Delta}),$$

where $k = \log q$, then

$$\Phi_\mu(x, y) = \sum q^{(\xi^2 + pr\eta^2)/\Delta} e^{2(\xi x + pr\eta y)i},$$

where the summation extends over all pairs of integers $\xi$, $\eta$ such that

$$\xi \equiv rs\mu, \quad \eta \equiv -t\mu \pmod{\Delta}.$$

Then

$$\Phi(x, y) \equiv \sum_{\mu=0}^{\Delta-1} \Phi_\mu(x, y) = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where

$$X = \Delta^{1/2}(x + y\rho\tau)$$

and the summation extends over all the numbers of $S$.

It is readily shown that

$$\Psi(x_1, y_1) \equiv \vartheta_3(rx_1, q^r)\vartheta_3(py_1, q^p) = \sum_{\Omega_1} q^{N(\Omega_1)} e^{2iR(\Omega_1 X_1')},$$

where

$$X_1 = x_1\rho + y_1\tau$$

and the summation extends over all the numbers of $S_1$.

Let $x_1$, $y_1$ be defined in terms of $x$, $y$ by

$$X_1 = XU.$$

By means of $\Omega_1 = \Omega U$ we have a one-to-one correspondence between the terms of $\Phi$ and $\Psi$. For corresponding terms,

$$N(\Omega_1) = N(\Omega)N(U) = N(\Omega),$$
$$\Omega_1 X_1' = (\Omega U)(U'X') = \Omega X'.$$

Hence corresponding terms are equal and

$$\Phi(x, y) = \Psi(x_1, y_1).$$

If this is expressed in terms of theta functions and $y$ is replaced by $-y/pr$, we obtain Schroeter's formula:

$$\vartheta_3(rsx + ty, q^r)\vartheta_3(-ptx + sy, q^p) = \sum_{\mu=0}^{\Delta-1} q^{r\mu^2} e^{2(rsx+ty)\mu i}\vartheta_3(\Delta x - rs\mu ki, q^\Delta)\vartheta_3(\Delta y - prt\mu ki, q^{pr\Delta}).$$

8. **An analogue of Schroeter's identity.** Let $p$, $r$, $s$, $t$ be defined as in the preceding paragraph, except that

$$rs^2 + pt^2 = \Delta^2,$$

where $\Delta$ is a positive integer. Let $S$ $[S_1]$ be the totality of numbers in the form

$$\Omega = \xi + \eta\rho\tau \quad [\Omega_1 = \xi_1\rho + \eta_1\tau],$$

where $\xi$, $\eta$ $[\xi_1, \eta_1]$ are integers such that

$$t\xi \equiv -rs\eta \qquad [t\xi_1 \equiv s\eta_1] \pmod{\Delta}.$$

Let $U = s\rho + t\tau$. If

$$\Omega U = \Delta\Omega_1, \tag{9}$$

it may be shown that $\Omega_1$ belongs to $S_1$ if and only if $\Omega$ belongs to $S$. Let $\mu$ be an arbitrarily chosen integer and

$$\Phi_\mu(x, y) \equiv q^{r\mu^2\Delta^2} e^{2r(sx-pty)\mu i}\vartheta_3(\Delta x - rs\mu\Delta ki, q^{\Delta^2})\vartheta_3(pr\Delta y + prt\mu\Delta ki, q^{pr\Delta^2}).$$

Then

$$\Phi_\mu(x, y) = \sum q^{\xi^2 + pr\eta^2} e^{2(\xi x + pr\eta y)i},$$

where the summation extends over all pairs of integers $\xi$, $\eta$ such that

$$\xi \equiv rs\mu, \quad \eta \equiv -t\mu \pmod{\Delta}.$$

Then

$$\Phi(x, y) \equiv \sum_{\mu=0}^{\Delta-1} \Phi_\mu(x, y) = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where the summation extends over all the numbers of $S$ and $X = x + y\rho\tau$.

If we set

$$\Psi(x_1, y_1) \equiv \sum_{\mu=0}^{\Delta-1} q^{\Delta^2\mu^2} e^{2\mu(rsx_1+pty_1)i} \vartheta_3(r\Delta x_1 - rs\mu\Delta ki, q^{r\Delta^2}) \vartheta_3(p\Delta y_1 - pt\mu\Delta ki, q^{p\Delta^2}),$$

it may be shown in a similar manner that

$$\Psi(x_1, y_1) = \sum_{\Omega_1} q^{N(\Omega_1)} e^{2iR(\Omega_1 X_1')},$$

where the summation extends over all the numbers of $S_1$ and $X_1 = x_1\rho + y_1\tau$.

By (9), the terms of $\Phi$ are in one-to-one correspondence with the terms of $\Psi$; and furthermore, for corresponding terms, $N(\Omega) = N(\Omega_1)$. Hence if we set

$$\Delta X_1 = XU$$

it follows that

$$\Phi(x, y) = \Psi(x_1, y_1).$$

If we express this in terms of theta functions and replace $q^{\Delta^2}, \Delta x, pr\Delta y, r\Delta x_1, p\Delta y_1$ by $q, x, y, x_1, y_1$ respectively, we obtain the identity below.

THEOREM 2. *If $p, r$ are positive integers and $s, t, u, v$ are integers such that*

$$rs^2 + pt^2 = \Delta^2,$$

*where $\Delta$ is a positive integer prime to $t$, then*

$$\sum_{\mu=0}^{\Delta-1} q^{r\mu^2} e^{2\mu x_1 i} \vartheta_3\left(x - \frac{rs\mu ki}{\Delta}, q\right) \vartheta_3\left(y + \frac{prt\mu ki}{\Delta}, q^{pr}\right)$$

$$= \sum_{\mu=0}^{\Delta-1} q^{\mu^2} e^{2\mu x i} \vartheta_3\left(x_1 - \frac{rs\mu ki}{\Delta}, q^r\right) \vartheta_3\left(y_1 - \frac{pt\mu ki}{\Delta}, q^p\right),$$

*where*

$$x_1 = (rsx - ty)/\Delta, \quad y_1 = (ptx + sy)/\Delta.$$

For the case $r = s = t = 1$, $p = 3$, the identity of the above theorem is equivalent to the identity mentioned at the end of §6. For $r = p = 2$, $s = t = 1$, the above identity reduces to

$$\vartheta_3(x, q)\vartheta_3(y, q) = \vartheta_3(x_1, q^2)\vartheta_3(y_1, q^2) + \vartheta_2(x_1, q^2)\vartheta_2(y_1, q^2),$$
$$x_1 = x - y, \qquad y_1 = x + y.$$

This is a special case of Schroeter's identity.

9. **An extension of Schroeter's identity.** In this and the next paragraph we shall obtain identities in functions of four variables similar to Schroeter's identity and to the identity of the preceding paragraph.

Let $a, b, c$ be positive integers and let $s, t, u, v$ be integers such that

$$s^2 + bct^2 + cau^2 + abv^2 \equiv \Delta$$

is prime to $\delta \equiv s^2 + bct^2$. Let

$$I = (bc)^{1/2}i, \quad J = (ca)^{1/2}j, \quad K = (ab)^{1/2}k,$$

where $i, j, k$ are the quaternion units, and let

$$A = su - btv, \qquad B = sv + ctu.$$

Let $S$ be the totality of quaternions in the form

$$\Omega = \Delta^{-1/2}(\xi + \eta I + \zeta J + \lambda K),$$

where $\xi, \eta, \zeta, \lambda$ are integers such that

$$\begin{aligned} \delta\xi &\equiv acA\zeta + abB\lambda, \\ \delta\eta &\equiv -aB\zeta + aA\lambda \end{aligned} \pmod{\Delta}. \tag{10}$$

Let $S_1$ be the totality of quaternions in the form

$$\Omega_1 = \xi_1 + \eta_1 I + \zeta_1 J + \lambda_1 K,$$

where $\xi_1, \eta_1, \zeta_1, \lambda_1$ are integers, and let

$$U = \Delta^{-1/2}(s + tI + uJ + vK).$$

It may be shown that if

$$\Omega U = \Omega_1, \tag{11}$$

then $\Omega_1$ belongs to $S_1$ if and only if $\Omega$ belongs to $S$. We have therefore by means of (11) a one-to-one correspondence between the elements of $S$ and those of $S_1$.

If $\xi, \eta, \zeta, \lambda$ satisfy (10), there is a uniquely determined pair of integers $\mu, \nu$; $0 \leqq \mu < \Delta$, $0 \leqq \nu < \Delta$, such that

$$\xi \equiv \delta\mu, \quad \eta \equiv \delta\nu \pmod{\Delta}. \tag{12}$$

By means of (10) we find

$$\zeta \equiv -A\mu + bB\nu, \quad \lambda \equiv -(B\mu + cA\nu) \pmod{\Delta}. \tag{13}$$

Conversely, if $\mu, \nu$ are arbitrarily chosen integers and if $\xi, \eta, \zeta, \lambda$ satisfy (12) and (13), they also satisfy (10). Let

$$\rho = -A\mu + bB\nu, \quad \tau = -(B\mu + cA\nu),$$

$$\Phi_{\mu,\nu} \equiv q^{\delta(\mu^2+bc\nu^2)} e^{2(\delta\mu x + bc\delta\nu y + ca\rho z + ab\tau w)i} \vartheta_3(\Delta x - \delta\mu ki, q^{\Delta})$$

$$\cdot \vartheta_3[bc(\Delta y - \delta\nu ki), q^{bc\Delta}] \vartheta_3[ca(\Delta z - \rho ki), q^{ca\Delta}] \vartheta_3[ab(\Delta w - \tau ki), q^{ab\Delta}]$$

where $k = \log q$. It may be shown that

$$\Phi_{\mu\nu} = \sum q^{(\xi^2+bc\eta^2+ac\zeta^2+ab\lambda^2)/\Delta} e^{2(\xi x + bc\eta y + ca\zeta z + ab\lambda w)i}$$

where the summation extends over all quadruples $\xi$, $\eta$, $\zeta$, $\lambda$ which satisfy (12) and (13).

If we set $X = \Delta^{1/2}(x + yI + zJ + wK)$ it follows that

$$\Phi(x, y, z, w) \equiv \sum_{\mu,\nu=0}^{\Delta-1} \Phi_{\mu\nu} = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where the summation extends over all the elements of $S$. It is readily shown that if

$$\Psi(x_1, y_1, z_1, w_1) \equiv \vartheta_3(x_1, q)\vartheta_3(bcy_1, q^{bc})\vartheta_3(caz_1, q^{ca})\vartheta_3(abw_1, q^{ab}),$$

then

$$\Psi(x_1, y_1, z_1, w_1) = \sum_{\Omega_1} q^{N(\Omega_1)} e^{2iR(\Omega_1 X_1')},$$

where the summation extends over all the elements of $S_1$ and $X_1 = x_1 + y_1 I + z_1 J + w_1 K$. By means of (11) there is a one-to-one correspondence between the terms of $\Phi$ and those of $\Psi$. If we let $x_1, y_1, z_1, w_1$ be defined in terms of $x, y, z, w$ by $X_1 = XU$, it may be shown as in the preceding proofs that the corresponding terms of $\Phi$ and $\Psi$ are equal and hence

$$\Phi(x, y, z, w) = \Psi(x_1, y_1, z_1, w_1).$$

Expressing this in terms of theta functions, replacing $y, z, w$ by $y/(bc)$, $z/(ca)$, $w/(ab)$ respectively, and making similar replacements for $y_1, z_1, w_1$, we obtain the identity below.

THEOREM 3. *Let $a, b, c$ be positive integers and let $s, t, u, v$ be integers such that*

$$s^2 + bct^2 + cau^2 + abv^2 \equiv \Delta$$

*is prime to $\delta = s^2 + bct^2$. Let $A = su - btv$, $B = sv + ctu$. Then*

$$\vartheta_3(x_1, q)\vartheta_3(y_1, q^{bc})\vartheta_3(z_1, q^{ca})\vartheta_3(w_1, q^{ab}) = \sum_{\mu,\nu=0}^{\Delta-1} q^{\delta(\mu^2+bc\nu^2)} e^{2(\delta\mu x + \delta\nu y + \rho z + \tau w)i}$$

$$\cdot \vartheta_3(\Delta x - \delta\mu ki, q^{\Delta})\vartheta_3(\Delta y - bc\delta\nu ki, q^{bc\Delta})\vartheta_3(\Delta z - ca\rho ki, q^{ca\Delta})\vartheta_3(\Delta w - ab\tau ki, q^{ab\Delta});$$

$$x_1 = sx - ty - uz - vw,$$
$$y_1 = bctx + sy + bvz - cuw,$$
$$z_1 = caux - avy + sz + ctw,$$
$$w_1 = abvx + auy - btz + sw,$$
$$\rho = -A\mu + bB\nu,$$
$$\tau = -B\mu - CA\nu.$$

10. **An extension of Theorem 2.** Let $a, b, c$ be positive integers and let $s, t, u, v$ be integers such that

$$s^2 + bct^2 + cau^2 + abv^2 = \Delta^2,$$

where $\Delta$ is a positive integer prime to $\delta = s^2 + bct^2$. Let $A, B, \rho, \tau$ be defined as in the preceding paragraph and let

$$A_1 = su + btv, \quad B_1 = sv - ctu.$$

Let $S$ be the totality of quaternions in the form

$$\Omega = \xi + \eta I + \zeta J + \lambda K,$$

where $\xi, \eta, \zeta, \lambda$ are integers which satisfy (10). Let $S_1$ be the totality of quaternions in the form

$$\Omega_1 = \xi_1 + \eta_1 I + \zeta_1 J + \lambda_1 K,$$

where $\xi_1, \eta_1, \zeta_1, \lambda_1$ are integers such that

$$\begin{aligned} \delta\xi_1 &\equiv -acA_1\zeta_1 - abB_1\lambda_1, \\ \delta\eta_1 &\equiv aB_1\zeta_1 - aA_1\lambda_1 \end{aligned} \pmod{\Delta}. \tag{14}$$

As in the preceding paragraph, if $\xi, \eta, \zeta, \lambda$ satisfy (10), there is a uniquely determined pair of integers $\mu, \nu$; $0 \leqq \mu < \Delta$, $0 \leqq \nu < \Delta$, such that

$$\xi \equiv \delta\mu, \quad \eta \equiv \delta\nu, \quad \zeta \equiv \rho, \quad \eta \equiv \tau \pmod{\Delta}. \tag{15}$$

If we set

$$\Phi_{\mu\nu} \equiv q^{\delta(\mu^2 + bc\nu^2)\Delta^2} e^{2(\delta\mu x + \delta bc\nu y + ca\rho z + ab\tau w)i} \vartheta_3(\Delta x - \delta\mu\Delta ki, q^{\Delta^2})$$
$$\cdot \vartheta_3(bc\Delta y - \delta\nu bc\Delta ki, q^{bc\Delta^2}) \vartheta_3(ca\Delta z - \rho ca\Delta ki, q^{ca\Delta^2}) \vartheta_3(ab\Delta w - \tau ab\Delta ki, q^{ab\Delta^2}),$$

it may be shown that

$$\Phi_{\mu\nu} = \sum q^{\xi^2 + bc\eta^2 + ca\zeta^2 + ab\lambda^2} e^{2(\xi x + bc\eta y + ca\zeta z + ab\lambda w)i},$$

where the summation extends over all quadruples which satisfy (15). Hence

$$\Phi(x, y, z, w) \equiv \sum_{\mu, \nu=0}^{\Delta-1} \Phi_{\mu\nu} = \sum_{\Omega} q^{N(\Omega)} e^{2iR(\Omega X')},$$

where the summation extends over all the elements of $S$ and $X=x+I+zJ+wK$.

Set

$$\rho_1 = A_1\mu - bB_1\nu, \quad \tau_1 = B_1\mu + cA_1\nu.$$

In $\Phi_{\mu\nu}$ replace $\rho, \tau$ by $\rho_1, \tau_1$ respectively and let $\Psi_{\mu\nu}$ be the resulting function. It may be shown as for $\Phi$ that

$$\Psi(x_1, y_1, z_1, w_1) = \sum_{\mu,\nu=0}^{\Delta-1} \Psi_{\mu\nu} = \sum_{\Omega_1} q^{N(\Omega_1)} e^{2\pi i R(\Omega_1 X_1')},$$

where the summation extends over all the elements of $S_1$ and $X_1=x_1+y_1I+z_1J+w_1K$. If we set $\Delta X_1=XU$, $\Delta\Omega_1=\Omega U$ where $U=s+tI+uJ+wK$, it may be shown that $\Omega_1$ belongs to $S_1$ if and only if $\Omega$ belongs to $S$ and that the corresponding terms of $\Phi$ and $\Psi$ are equal. Hence

$$\Phi(x, y, z, w) = \Psi(x_1, y_1, z_1, w_1).$$

If we express this in terms of theta functions, replace $q^{\Delta^2}$, $\Delta x$, $\Delta bcy$, $\Delta caz$, $\Delta abw$ by $q$, $x$, $y$, $z$, $w$ respectively and make similar replacements for $x_1, y_1, z_1, w_1$, we obtain the identity below.

THEOREM 4. *Let $a, b, c$ be positive integers and let $s, t, u, v$ be integers such that*

$$s^2 + bct^2 + cau^2 + abv^2 = \Delta^2,$$

*where $\Delta$ is a positive integer prime to $\delta \equiv s^2+bct^2$. Let*

$$A = su - btv, \quad A_1 = su + btv, \quad \rho = -A\mu + bB\nu, \quad \rho_1 = A_1\mu - bB_1\nu,$$
$$B = sv + ctu, \quad B_1 = sv - ctu, \quad \tau = -B\mu - cA\nu, \quad \tau_1 = B_1\mu + cA_1\nu,$$

*where $\mu, \nu$ are arbitrarily chosen integers. Let*

$$F(x, y, z, w) \equiv \sum_{\mu,\nu=0}^{\Delta-1} q^{\delta(\mu^2+bc\nu^2)} e^{2(\delta\mu x+\delta\nu y+\rho z+\tau w)i/\Delta} \vartheta_3\left(x - \frac{\delta\mu ki}{\Delta}, q\right)$$
$$\cdot \vartheta_3\left(y - \frac{\delta bc\nu ki}{\Delta}, q^{bc}\right) \vartheta_3\left(z - \frac{\rho cak i}{\Delta}, q^{ca}\right) \vartheta_3\left(w - \frac{\tau abk i}{\Delta}, q^{ab}\right),$$

*and let $F_1$ be the function obtained from $F$ by replacing $\rho, \tau$ by $\rho_1, \tau_1$ respectively. Then*

$$F(x, y, z, w) = F_1(x_1, y_1, z_1, w_1)$$

*where*

$$x_1 = (sx - ty - uz - vw)/\Delta, \qquad z_1 = (caux - avy + sz + ctw)/\Delta,$$
$$y_1 = (bctx + sy + bvz - cuw)/\Delta, \qquad w_1 = (abvx + auy - btz + sw)/\Delta.$$

UNIVERSITY OF KENTUCKY,
LEXINGTON, KY.

# THE CONSISTENCY AND ULTIMATE DISTRIBUTION OF OPTIMUM STATISTICS*

BY
HAROLD HOTELLING

1. **Definitions.** It has been customary to assume somewhat loosely that when a quantity is calculated from a random sample to estimate a parameter of a hypothetical frequency distribution, the accuracy of the determination will increase without limit as the number in the sample increases. A *consistent statistic* is a function of the observations actually possessing this property, which we define more precisely as follows. If $p$ is the true value of the parameter (e.g. the true declination of a star) and $p'$ an estimate of $p$ calculated from data (e.g. the geometric mean of the declinations arrived at by a class in astronomy), then $p'$ is a consistent statistic if for every two positive numbers $\delta$ and $\epsilon$ a number $N$ exists such that if the sample contains more than $N$ individuals the probability is less than $\epsilon$ that

$$|p - p'| > \delta.$$

However, some statistics, such as averages of biased measurements, or rank correlation coefficients as direct estimates of product-moment correlations, are not consistent. The utility of such a statistic, if any, lies in ease of computation, together with sufficient accuracy for ordinary purposes, though the accuracy cannot be increased beyond a certain point by multiplying observations.

It has also been common to assume loosely that the probability distribution of a statistic calculated from $n$ observations has a shape which approaches that of the normal ("Gaussian") curve as $n$ increases. But in some cases this is not true. For example the range $r$ of a sample of $n$ from a rectangular population of breadth $p$ has the probability

$$\frac{n(n-1)r^{n-2}(p-r)dr}{p^n} \qquad (r < p)$$

of falling in the range $dr$. The form of this distribution does not approach normality.

If the probability of obtaining an observation in the range $x \pm \frac{1}{2}dx$ is $f(x, p)dx$, and if observations $x_1, \cdots, x_n$ have been obtained, then

* Presented to the Society, December 30, 1929; received by the editors in December, 1929.

$$L = \sum_{i=1}^{n} \log f(x_i, p),$$

or any quantity whose difference from this does not involve $p$, has been defined by R. A. Fisher as the logarithm of the *likelihood*.* The same definition holds if the probability that an observation shall be exactly $x$ is finite and equal to $f(x, p)$. Thus we deal simultaneously with continuous and with discontinuous distributions of $x$. For a multivariate distribution of any kind and for a multiplicity of parameters this and the following definitions may be extended in obvious fashion.

The value $\hat{p}$ of $p$ which makes $L$ a maximum is a function of $x_1, \cdots, x_n$; it will be convenient to use Fisher's designation of this function as the *optimum statistic* or *optimum estimate of* $p$. Examples of optimum statistics are the mean of a sample from a normal distribution for estimating the mean of the population, and the range $r$ of a sample from a rectangular distribution for estimating the range of this distribution. The *variance* of any quantity, the square of its standard error, is defined as the mean value, or mathematical expectation, of the square of its deviation from its mean value. The *covariance* of two quantities is the mean value of the product of their deviations from their respective mean values. In the paper cited Fisher shows† that, if an optimum statistic has a distribution approaching the normal form, its variance approximates to the reciprocal of the mean value of $-\partial^2 L/\partial p^2$ as $n$ increases. As an extension to the case of simultaneous estimation of parameters $p_1, \cdots, p_k$, he uses for the variances and covariances of the optimum statistics the elements of the inverse of the matrix

$$\left|\left| -E\left(\frac{\partial^2 L}{\partial p_i \partial p_j}\right) \right|\right|$$

where $E$ signifies the mathematical expectation. In this way Fisher arrives at results of wide practical importance, including a condemnation of the general use of the method of moments for fitting frequency curves.

Apart from the question whether the hypothesis of an ultimately normal distribution is satisfied, it is not clear what conditions, particularly of continuity, are necessary in order that the proofs which have been given shall

---

* *On the mathematical foundations of theoretical statistics*, Philosophical Transactions, vol. 222A (1922), pp. 309–368.

† By a method similar to one used for the same purpose by Pearson and Filon, Philosophical Transactions, vol. 191A (1898), p. 229.

be valid. Thus Fisher's application (op. cit. p. 333) to the location of the Type III curve breaks down if $p \leqq 1$, though all values greater than $-1$ are possible.

We shall put $\phi = \log f$. Then since the total frequency of the population is unity, $\int e^{\phi} dx = 1$. Differentiating twice we have

$$n \int_{-\infty}^{\infty} (\partial\phi/\partial p)^2 e^{\phi} dx = - n \int_{-\infty}^{\infty} (\partial^2\phi/\partial p^2) e^{\phi} dx,$$

provided the second derivative exists and the integrals converge. The right-hand side of the equation is the mean value used by Fisher and others. Likewise for problems of simultaneous estimation of two or more parameters we find that the mean value of

$$\frac{\partial\phi}{\partial p_i} \frac{\partial\phi}{\partial p_j}$$

and that of

$$- \frac{\partial^2\phi}{\partial p_i \partial p_j}$$

are identical.

2. **Theorems.** We shall prove explicitly the theorems below for frequency functions of one continuous variable. However, the extensions to any number of variables are perfectly obvious; and the corresponding theorems for discrete variables follow immediately by replacing each value of the variable by an interval within which the probability may be supposed uniformly distributed.

The true value of $p$ will be denoted by $p_0$, its optimum estimate by $\hat{p}$, and the probability that a random value of $x$ lies between $\alpha$ and $\beta$ by $\int_{\alpha}^{\beta} f(x, p) dx$. The following properties of the distribution function $f$ will be used:

(a) $f$ is a continuous function of $x$ except on a set of values of $x$ of measure zero.

(b) $f$ is a continuous monotonic function of $p$ in a $p$-interval including $p_0$ for all values of $x$ in some interval.

(c) In a $p$-interval including $p_0$, $\partial f/\partial p$ is a continuous function of $p$ for every value of $x$; $x^2 \partial f/\partial p$ approaches a continuous function of $p$ as $x \to \pm\infty$; in some $x$-interval $\partial f/\partial p$ does not vanish.

The theorems are the following:

I. *If* (a) *and* (b) *are satisfied, $\hat{p}$ is a consistent statistic.*

II. *The distribution of $\hat{p}$ approaches the normal form if* (a) *and* (c) *are satisfied.*

III. *Under hypotheses* (a) *and* (c) *the variance of* $\hat{p}$ *bears to*

$$E\left(\frac{\partial L}{\partial p}\right)^2_{p=p_0} = n\int_{-\infty}^{\infty}[\partial \log f(x, p_0)/\partial p_0]^2 f(x, p_0)dx$$

$$= n\int_{-\infty}^{\infty}[\partial\phi(x, p_0)/\partial p_0]^2 e^{\phi(x,p_0)}dx$$

$$= -n\int_{-\infty}^{\infty}[\partial^2\phi(x, p_0)/\partial p_0^2]e^{\phi(x,p_0)}dx,$$

*a ratio which approaches unity as n increases.*

IV. *If* $f(x, p_1, \cdots, p_k)$ *is a frequency function satisfying the conditions* (a) *and* (c) *for each* $p_i$, *the optimum joint estimates of* $p_1, \cdots, p_k$ *have a distribution approaching the normal form as n increases. The variances and covariances, each multiplied by n, approach the elements of a matrix whose inverse consists of the elements*

$$\int_{-\infty}^{\infty}\frac{\partial\phi}{\partial p_i}\frac{\partial\phi}{\partial p_j}e^{\phi}dx,$$

*where* $\phi = \log f$.

In §6 we shall correct the usual proof of the formula for the covariance of two frequencies.

The hypothesis of Theorem II which is violated in locating a Type III curve for which the parameter of shape, $p$, is not greater than unity is that of continuity of the first derivative with respect to the parameter of location, which Fisher denotes by $m$. Variation of $m$ determines a translation of the curve along the axis. When $p=1$ the curve makes an angle at one end with the axis; when $p<1$ the angle is $\pi/2$. Translation therefore causes an ordinate corresponding to a fixed value of $x$ to change at a discontinuous rate when $p$ has such values. By a transformation of $x$ the curve can be thrown into a form tangent to the axis and so satisfying the hypotheses (c), but the transformation must depend upon the location as well as the shape of the curve.

3. **Reduction of infinite to finite range.** Three kinds of infinity are involved in the problem. The range of $x$ may be infinite, or may increase indefinitely as $p$ approaches some value. Secondly, we shall divide the range into subintervals whose lengths will eventually approach zero. Finally, the number in the sample is to increase indefinitely.

The infinite range we dispose of at once by the simple expedient of transforming the variable into one of finite range and using the invariantive property of the optimum statistic. Thus we may put $x=\tan\theta$. The distri-

bution $f(x, p)dx$ then becomes

$$F(\theta, p)d\theta = f(\tan\theta, p)\sec^2\theta d\theta,$$

and the likelihood,

$$L = \sum \log [f(\tan\theta_i, p)\sec^2\theta_i] = Q + \sum \log f(x_i, p),$$

where $x_i = \tan\theta_i$. Since $Q$ does not involve $p$, the value of $p$ which makes $L$ a maximum is exactly the same as before. However the range of the variable $\theta$ is finite.

The new frequency function satisfies the same conditions as the old concerning the continuity and non-vanishing of its first derivative with respect to $p$. This is true even at the ends of the range, since there

$$\partial F/\partial p = \lim (1 + x^2)\partial f/\partial p,$$

which is continuous by hypothesis (c). The integrals in Theorems III and IV have the same values whether calculated from the distribution of $x$ or from that of $\theta$.

4. **Approximation by grouped distribution.** Let the range of the variable $x$, which we shall from now on assume finite, be divided into $m$ intervals $J_1, J_2, \cdots, J_m$ of respective lengths $l_1, \cdots, l_m$. For convenience we take these intervals such that the frequencies

$$f_t(p) = \int_{J_t} f(x, p)dx \qquad (t = 1, 2, \cdots, m)$$

are all equal when $p$ has the value $p_0$ which it takes in the population in question. The maximum length $l$ can be made arbitrarily small by using a sufficiently great number $m$ of intervals, provided that in measuring the length of an interval we exclude any subintervals in which $\int f dx$ is zero.

Hypothesis (a) shows that if $x$ is in $J_t$, the ratio $f_t(p)/l_t$ can by increase of $m$ be made to differ as little as we please from $f(x, p)$, unless $x$ is one of a set of values of measure zero and so can be neglected in considering probabilities. Consequently, if in a sample of $n$ the number falling in the class $J_t$ be denoted by $n_t$, the logarithm of the likelihood derived from the grouped data, which may be taken as

$$L' = \sum_{t=1}^{m} n_t \log f_t(p) - \sum n_t \log l_t,$$

the second sum not involving $p$ and so not affecting the maximizing value, differs arbitrarily little from

$$L = \sum_{i=1}^{n} \log f(x_i, p).$$

This is true for every value of $p$; and by hypothesis $f(x, p)$, and therefore $L$ and $L'$, are continuous functions of $p$. Therefore if $\hat{p}'$ and $\hat{p}$ denote respectively the values maximizing $L'$ and $L$, the difference $|\hat{p}'-\hat{p}|$ can be made arbitrarily small by increasing $m$. Consequently if our theorems are true for data grouped in this way, and for every sufficiently large value of $m$, they are true also for ungrouped data. This fact enables us to concentrate our attention upon the variables $n_1, \cdots, n_m$, which are finite in number.

5. **Proof of Theorem I.** Consider the expression

$$L = \sum_{t=1}^{m} n_t \log z_t,$$

where $n_t$ is the number of observations in a particular sample of $n$ falling in the interval $J_t$, so that $\sum n_t = n$. Let

$$L = L_R \text{ when } z_t = n_t/n;$$
$$L = L_Z \text{ when } z_t = f_t(\hat{p}');$$
$$L = L_M \text{ when } z_t = f_t(p_0) = 1/m \qquad (t = 1, 2, \cdots, m).$$

Then $L_R$ is the maximum value of $L$ when the $z_t$ are subject only to the restriction

$$\sum_{t=1}^{m} z_t = 1; \tag{5.1}$$

$L_Z$ is the maximum value when they are subject to the severer condition that a value of $p$ exists such that

$$z_t = f_t(p) \qquad (t = 1, 2, \cdots, m) \tag{5.2}$$

and $L_M$ satisfies the last condition but is not the maximum. Hence

$$L_R \geqq L_Z \geqq L_M.$$

Now Bernoulli's law of great numbers shows that, when $n$ is large enough, we can assert with a probability greater than $1-\epsilon$ that

$$| n_t/n - f_t(p_0) | < \delta, \tag{5.3}$$

$\epsilon$ and $\delta$ being arbitrarily small. When this condition is satisfied, since $L$ is a continuous function of the $z_t$, $L_R-L_M$ has an upper limit which approaches zero with $\delta$. The same will therefore be true of $L_R-L_Z$. We shall next prove it true of each of the quantities

$$| f_t(\hat{p}') - f_t(p_0) | .$$

Since the hypotheses of Theorem I show that some at least of the functions

$f_i(p)$ have single-valued continuous inverses when $m$ is large enough, it will then follow that $|\hat{p}'-p_0|$, and so, by §4, $|\hat{p}-p_0|$, are arbitrarily small, thus establishing the theorem.

To fill in the gap in the argument, let $L_1$ be the maximum value of $L$ for a fixed value of $z_1$, and subject otherwise only to (5.1). Then $L_R \geqq L_1$. It is easy to see that this maximum is determined by

$$z_i = n_i(1 - z_1)/(n - n_1) \qquad (i = 2, 3, \cdots, m), \tag{5.4}$$

and equals

$$L_1 = n_1 \log z_1 + (n - n_1)[\log (1 - z_1) - \log (n - n_1)] + \sum_{i=2}^{m} n_i \log n_i.$$

Obviously $L_1$ is a continuous function of $z_1$ from 0 to 1, not inclusive, and

$$dL_1/dz_1 = n_1/z_1 - (n - n_1)/(1 - z_1)$$

changes sign only for $z_1 = n_1/n$. By the theorem on inverse functions, $z_1$ is therefore a double-valued *continuous* function of $L_1$, the two branches becoming equal for $z_1 = n_1/n$, where $L_1 = L_R$. Hence, for any $\delta' > 0$,

$$|z_1 - n_1/n| < \delta',$$

if $L_R - L_1$ is sufficiently small. This will be true a fortiori if $z_2, z_3, \cdots, z_m$ have values other than (5.4) and if $L_R - L$ is sufficiently small, for $L_R - L_1$ is still smaller.

In the same way each of the quantities

$$|z_i - n_i/n| < \delta'$$

when $L_R - L$ is small enough. Putting $L = L_Z$, which as we have seen will be arbitrarily close to $L_R$, we have

$$|f_i(\hat{p}') - n_i/n| < \delta'$$

and therefore from (5.3),

$$|f_i(\hat{p}') - f_i(p_0)| < \delta + \delta',$$

the required inequality.

A geometrical interpretation may render the foregoing argument more intelligible. The $z_i$ being coördinates in $m$-space, (5.1) is a hyperplane containing the curve (5.2), the point $R$ representing the observations, the point $Z$ on the curve at which the parameter $p$ takes the optimum value $p'$, and the point $M$ where $p$ takes its true value $p_0$. The likelihood $L$ is constant over

a system of approximately spherical hypersurfaces about $R$. The point $Z$ is the point of the curve which lies on the smallest of the approximate spheres meeting the curve, and is therefore approximately the nearest point on the curve to $R$. The argument shows with probability greater than $1-\epsilon$ that $Z$ lies arbitrarily close to $R$ and therefore to $M$, and that the values of the parameter at $Z$ and at $M$ differ arbitrarily little.

The curve (5.2) may be quite irregular. For example the arc length may fail to exist, owing to infinitely rapid oscillations of some of the ordinates of the frequency function when $p$ varies. The conditions of Theorem I require only one coördinate—that is, the integral of the frequency function in some $x$-interval—to be continuous and monotonic with respect to $p$, and then only in a short $p$-interval. The subsequent theorems require functions of a smoother sort.

6. **Distribution of class frequencies.** The familiar demonstration by means of Stirling's formula shows that the probability that $n_t$ will fall between any limits is given by the integral between these limits of a normal curve of variance proportional to $n$, apart from terms which vanish as $n$ increases. We now examine the joint normal distribution of all these class frequencies in order to deduce that of $\hat{p}$.

Let the deviations

$$r_t = n_t/n - 1/m$$

of the observed relative frequencies from their values in the population be taken as the cartesian coördinates of a point $R$ in $m$ dimensions. The values of the $n_t/n$ obtained from random samples have the binomial distribution with mean $1/m$, and therefore with variance

$$\sigma^2 = (1 - 1/m)/(mn).$$

If each $r_t$ were determined from a different random sample of $n$, the points $R$ would form a globular cluster in the $m$-space. The distribution of probability would have, in fact, a spherical symmetry whenever $n$ is large enough for the binomial to be replaced by the normal distribution; for the probability of $R$ falling in the volume element $dr_1 dr_2 \cdots dr_m$ would be the product of the $m$ separate probabilities, or

$$e^{-U/2\sigma^2} dr_1 dr_2 \cdots dr_m/(2\pi\sigma^2)^{m/2},$$

where

$$U = \sum_{t=1}^{m} r_t^2.$$

However the points $R$ which we are considering are confined to the hyperplane $V_{m-1}$ of equation $\sum r_t = 0$, since all the coördinates are derived from

the same sample. The imposition of this restriction destroys the mutual independence of the class frequencies, giving instead a covariance between any two of them which will now be seen to equal $-1/(nm^2)$.

The derivations which have been given of the covariance or correlation between sampling deviations in class frequencies are objectionable because of the inconsistent assumption that the deviation of an observed class frequency from its mean value is compensated, in the same sample, by deviations of the opposite sign in all the other class frequencies distributed exactly in proportion to the mean values. Since we are attempting an accurate treatment we must therefore turn aside to give an accurate proof of this much-used proposition. The accurate proof is simpler than the customary one.

In any problem of sampling grouped values, let $f_i$ be the mean frequency in the $i$th class in a sample of $n$, and let $f_i+\delta f_i$ be the observed frequency. Then from the elementary theory of the Bernoulli distribution,

$$E(\delta f_i)^2 = f_i(1 - f_i/n),$$
$$E(\delta f_j)^2 = f_j(1 - f_j/n).$$

Combining the $i$th and $j$th classes we have, by the same principle,

$$E(\delta f_i + \delta f_j)^2 = (f_i + f_j)\left(1 - \frac{f_i + f_j}{n}\right) \qquad (i \neq j).$$

Subtracting the first two of these equations from the last and dividing by 2 we have the familiar expression for the covariance,

$$r_{f_i f_j}\sigma_{f_i}\sigma_{f_j} = E(\delta f_i \delta f_j) = -f_i f_j/n.$$

Substituting $f_i = f_j = 1/m$, we have the result needed for present purposes: the covariance of any two of our $r_i$ is $-1/(nm^2)$. The density of the points $R$ in $V_{m-1}$ is therefore proportional to

$$\exp\left(-\frac{1}{2}\sum_{s=1}^{m-1}\sum_{t=1}^{m-1} a_{st} r_s r_t\right),$$

where $||a_{st}||$ is the inverse of the matrix of variances and covariances. In accordance with the results just obtained, the latter matrix is

$$\begin{Vmatrix} (1-1/m)/(nm) & -1/(nm^2) & \cdots & -1/(nm^2) \\ -1/(nm^2) & (1-1/m)/(nm) & \cdots & -1/(nm^2) \\ \cdots & \cdots & \cdots & \cdots \\ -1/(nm^2) & -1/(nm^2) & \cdots & (1-1/m)/(nm) \end{Vmatrix}.$$

It is easy by straightforward algebra to obtain the inverse of this matrix or of the more general one in which the class frequencies are unequal, though Karl Pearson resorted for the latter purpose to a complicated trigonometric method.* With the help of $\sum_{t=1}^{m} r_t = 0$ we obtain simply

$$\sum_{s=1}^{m-1}\sum_{t=1}^{m-1} a_{st} r_s r_t = mn \sum_{t=1}^{m} r_t^2 .$$

Thus the density distribution in $V_{m-1}$ is spherically symmetrical.

Let us take new cartesian coördinates $y_1, y_2, \cdots, y_m$ with the same origin as the $r_t$ and such that $y_m = 0$ on $V_{m-1}$. On this hyperplane the density distribution, being symmetrical, will be given by

$$(mn/(2\pi))^{(m-1)/2} \exp\left( -\frac{1}{2} mn \sum_{t=1}^{m-1} y_t^2 \right) dy_1 \cdots dy_{m-1}.$$

Now suppose that each point $R$ is projected orthogonally upon a line which, without loss of generality, we assume to be the $y_1$-axis. The linear density of the resulting points is found by integrating the last expression with respect to $y_2, y_3, \cdots, y_{m-1}$, an easy matter because it is the product of factors involving only one variable each. The result shows that the linear density of the projected points is normally distributed with variance

$$\sigma_z^2 = 1/(mn). \tag{6.1}$$

7. **Proof of normal distribution.** In addition to $R$, each sample determines a point $Z$ in $V_{m-1}$, the coördinates $z_1, \cdots, z_m$ of $Z$ being the deviations from $1/m$ of the proportions falling into the several classes of the total frequency in a population having for its parameter $p$ the value $\hat{p}'$ for which the likelihood of the given sample is a maximum. We shall now find an approximate geometrical construction for $Z$ when $R$ is known.

Since the coördinates of $Z$ are the relative frequencies in a population of the form under consideration, minus $1/m$, they must satisfy the equations

$$z_t = f_t(p) - 1/m \qquad (t = 1, 2, \cdots, m) \tag{7.1}$$

which define a curve in $V_{m-1}$ in terms of the parameter $p$. This curve $C$ passes through the center $P$ of our globular cluster; and since it follows from the hypotheses of Theorem II that the functions $f_t(p)$ have continuous first derivatives of which one at least does not vanish, $C$ is approximated in a neighborhood of $P$ by the tangent straight line, along which the distance measured from $P$ is

* *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philosophical Magazine, vol. 50 (1900), p. 157.

$$s = (p - p_0)\{\sum[f_i'(p_0)]^2\}^{1/2}. \tag{7.2}$$

The point $Z$ for a particular sample is that for which

$$L = \sum_{i=1}^{m} n_i \log (z_i + 1/m) \tag{7.3}$$

is a maximum for variations of $z_i$ subject to (7.1). For any constant value of $L$, (7.3) is the equation of a hypersurface. The problem of maximizing $L$ subject only to the condition

$$\sum z_i = 0 \tag{7.4}$$

would be the problem of determining one of these hypersurfaces tangent to the hyperplane (7.4). The solution of this problem is simply $z_i = n_i/n - 1/m$; the point of tangency is $R$.

Putting $L_R$ for the value of $L$ at this point, $L_Z$ for its value at $Z$, and $\delta z_1, \cdots, \delta z_m$ for the differences in coördinates of these points, we have, apart from terms of higher order,

$$L_Z = L_R + \sum_i (\partial L/\partial z_i)_R \, \delta z_i + \tfrac{1}{2} \sum_s \sum_t (\partial^2 L/\partial z_s \partial z_t)_R \, \delta z_s \, \delta z_t.$$

Now from (7.3) we have at $R$, where $z_i = n_i/n - 1/m$,

$$\partial L/\partial z_i = n, \qquad \partial^2 L/\partial z_s \, \partial z_t = 0 \text{ if } s \neq t, \qquad = -n^2/n_i \text{ if } s = t.$$

Moreover $\sum \delta z_i = 0$, since both $Z$ and $R$ lie in $V_{m-1}$. Thus $Z$ is the point on the curve $C$ for which

$$\sum (\delta z_i)^2/n_i$$

is a minimum. For large values of $n$ the denominators tend to equality and the point $Z$ therefore to the nearest point on the curve to $R$.

This shows that the points $R$ are in the limit projected orthogonally upon $C$, or approximately upon the tangent line. The distance $s$ along this line will then have a normal distribution with variance (6.1). The distribution of $p$ will therefore approach the normal form with a variance, derived from that of $s$ by means of (7.2), equal to

$$1/\{mn \sum [f_i'(p_0)]^2\}. \tag{7.5}$$

Since this $p$ is identical with $\hat{p}'$, which approximates $\hat{p}$ for the fine grouping implied by large values of $m$, Theorem II is established.

It is important that the last of the conditions (c) prevents the vanishing of the denominator of (7.5). If this condition were not insisted upon it would be easy to find parameters whose optimum estimates did not tend to a normal distribution and whose sampling deviations might therefore easily

be misinterpreted. Such for example would be $q=(p-p_0)^2$; the distribution of an optimum estimate of $q$ would approach the form $e^{-kq}\,dq/[2(k\pi q)^{1/2}]$.

8. **Evaluation of variance.** To prove Theorem III we multiply (7.5) by $n$ and let $m$ increase without limit. Since by definition

$$f_i(p_0) = 1/m,$$

and since, apart from terms of higher order in $m^{-1}$,

$$f_i(p) = f(x, p)l_i,$$

so that

$$f_i'(p)^2/f_i(p) = [\partial \log f(x, p)/\partial p]^2 f(x, p)l_i,$$

it follows that the expression

$$m\sum[f_i'(p_0)]^2 = \sum[f_i'(p_0)]^2/f_i(p_0)$$

will, as $m$ increases, approach the value for $p=p_0$ of

$$\int_{-\infty}^{\infty}(\partial \log f/\partial p)^2 f\,dx, \quad \text{or} \quad \int_{-\infty}^{\infty}(\partial\phi/\partial p)^2 e^{\phi}dx,$$

where $f=e^{\phi}$ is taken as zero outside the range of $x$. This proves Theorem III.

9. **Extension to more than one parameter.** For each of $k$ parameters $p_1, \cdots, p_k$ separately, Theorems I, II, and III hold. Since the optimum estimate of each has in the limit a normal distribution, their joint distribution approaches that for which the frequency element is proportional to $e^{-nT/2}$, where

$$T = \sum_{i=1}^{k}\sum_{j=1}^{k} g_{ij}(p_i - p_{i0})(p_j - p_{j0})$$

is a positive definite quadratic form. Here $p_{i0}$ represents the true value of $p_i$, and the carats have been dropped from the $p_i$ in denoting the optimum estimates. The matrix of the $g_{ij}$ is the inverse of that of the variances and covariances.

Let us apply a linear transformation which reduces $T$ to

$$T' = q_1^2 + q_2^2 + \cdots + q_n^2,$$

and let the inverse transformation be

$$q_h = \sum_i A_{hi}(p_i - p_{i0}). \tag{9.1}$$

Equating coefficients,

$$g_{ij} = \sum_h A_{hi}A_{hj}. \tag{9.2}$$

Taking the $q$'s instead of the $p$'s as parameters we observe from the form of $T'$ that their optimum estimates are in the limit distributed normally and independently with variance $1/n$. Hence from Theorem III,

$$\int (\partial\phi/\partial q_h)^2 e^\phi \, dx = 1. \tag{9.3}$$

To show further that

$$\int (\partial\phi/\partial q_h)(\partial\phi/\partial q_l) e^\phi \, dx = 0 \text{ if } h \neq l, \tag{9.4}$$

we put

$$q_h = (u_h - u_l)/2^{1/2},$$
$$q_l = (u_h + u_l)/2^{1/2},$$

that is,

$$u_h = (q_h + q_l)/2^{1/2},$$
$$u_l = (-q_h + q_l)/2^{1/2}.$$

The left member of (9.4) equals one-half of

$$\int (\partial\phi/\partial u_h)^2 e^\phi \, dx - \int (\partial\phi/\partial u_l)^2 e^\phi dx.$$

But these integrals are equal because they are respectively $n$ times the variances of $u_h$ and $u_l$, both of which have, by their definition, the same variance as the $q$'s.

Now since, by (9.1),

$$\partial\phi/\partial p_i = \sum(\partial\phi/\partial q_h)(\partial q_h/\partial p_i) = \sum A_{hi}\partial\phi/\partial q_h,$$

we have

$$(\partial\phi/\partial p_i)(\partial\phi/\partial p_j) = \sum_h \sum_l A_{hi} A_{lj} (\partial\phi/\partial q_h)(\partial\phi/\partial q_l).$$

Taking the mean values of both sides and using (9.3), (9.4), and then (9.2),

$$\int [(\partial\phi/\partial p_i)(\partial\phi/\partial p_j)] e^\phi \, dx = \sum A_{hi} A_{hj} = g_{ij}.$$

Thus Theorem IV is proved.

STANFORD UNIVERSITY,
STANFORD UNIVERSITY, CALIF.

# THE NOTION OF THE GREEN'S FUNCTION IN THE THEORY OF INTEGRO-DIFFERENTIAL EQUATIONS. II*

BY

J. D. TAMARKIN†

1. **Problem and hypotheses.** We are concerned here with an integro-differential boundary value problem (★) consisting of an integro-differential equation

$$(1.1)\qquad \begin{aligned} L(u) &\equiv u^{(n)} + P_1(x,\rho)u^{(n-1)} + \cdots + P_n(x,\rho)u \\ &= f(x) + \int_a^b h(x,\xi;\rho)u(\xi)d\xi \end{aligned}$$

together with $n$ linear boundary conditions

$$(1.2)\qquad L_i(u) \equiv A_i(u,\rho) + B_i(u,\rho) + \int_a^b \alpha_i(x;\rho)u(x)dx = 0;$$

$$(1.3)\qquad \begin{aligned} P_i(x,\rho) &\equiv \rho^i p_{i0}(x) + \rho^{i-1}p_{i1}(x) + \cdots + p_{ii}(x), \\ A_i(u,\rho) &\equiv \sum_{s=0}^{n}\rho^s A_i^{(s)}(u), \quad B_i(u,\rho) \equiv \sum_{s=0}^{n}\rho^s B_i^{(s)}(u), \\ \alpha_i(x;\rho) &\equiv \sum_{s=0}^{n}\rho^s\alpha_{is}(x) \end{aligned}$$

[I; 10, 13, 19]. The coefficients $\phi_j(x;\rho)$ of [I; 19] are assumed here to be identically zeros while the function $h(x,\xi;\rho)$ is defined by

$$(1.4)\qquad h(x,\xi;\rho) \equiv p_{n0}(x)\sum_{s=0}^{n}\rho^s h_s(x,\xi).$$

We assume that our problem (★) satisfies hypotheses (**A**) and (**B**) of [I; 10, 13] while hypothesis (**C**) of (I; 19] is replaced by

(**C**) i. The function $h_n(x,\xi)$ has partial derivatives of the first order

---

* Presented to the Society, December 28, 1926; received by the editors June 24, 1930.

† The present paper is a continuation of a paper published in these Transactions (vol. 29 (1927), pp. 755–800) under the same title. We shall refer by [I] to this paper as to notation and bibliography of the subject. See also Errata to [I] at the end of the present Note.

which are of bounded variation in $(x, \xi)$.* The function $h_{n-1}(x, \xi)$ is of bounded variation in $(x, \xi)$. The remaining functions $h_{n-2}(x, \xi), \cdots, h_0(x, \xi)$ are bounded and integrable on $(a, b)$ with respect either to $x$, or to $\xi$, or to $(x, \xi)$.

ii. The function $h_n(x, \xi)$ considered as a kernel of a Fredholm integral equation possesses a reciprocal $\mathfrak{h}_n(x, \xi)$ as defined by

$$h_n(x, \xi) + \mathfrak{h}_n(x, \xi) = (h_n \cdot \mathfrak{h}_n)(x, \xi) = (\mathfrak{h}_n \cdot h_n)(x, \xi). \dagger$$

2. **A lemma of the theory of integral equations.** The following well known lemma‡ will be very useful for our discussion.

LEMMA 1. *If a kernel $K(x, \xi)$ is equal to the sum of two other kernels,*

$$K(x, \xi) = K_1(x, \xi) + K_2(x, \xi) \tag{2.1}$$

*and if the reciprocals of the kernel $K_1(x, \xi)$ and of the kernel*

$$K_0(x, \xi) \equiv K_2(x, \xi) - (\mathfrak{K}_1 \cdot K_2)(x, \xi) \tag{2.2}$$

*exist, then the reciprocal of $K(x, \xi)$ exists and is given by*

$$\mathfrak{K}(x, \xi) = \mathfrak{K}_1(x, \xi) + \mathfrak{K}_0(x, \xi) - (\mathfrak{K}_0 \cdot \mathfrak{K}_1)(x, \xi). \S \tag{2.3}$$

Formula (2.3) is easily proved if we write the integral equation

$$y(x) = f(x) + K \cdot y(x)$$

in the form

$$y(x) = f(x) + K_2 \cdot y(x) + K_1 \cdot y(x),$$

whence, by definition of the reciprocal and by the hypothesis of Lemma 1,

$$\begin{aligned} y(x) &= f(x) + K_2 \cdot y(x) - \mathfrak{K}_1 \cdot \{f(x) + K_2 \cdot y(x)\} \\ &= f(x) - \mathfrak{K}_1 \cdot f(x) + K_0 \cdot y(x), \\ y(x) &= f(x) - \mathfrak{K}_1 \cdot f(x) - \mathfrak{K}_0 \cdot \{f(x) - \mathfrak{K}_1 \cdot f(x)\}, \end{aligned}$$

which, being compared with

* A function $f(x, \xi)$ will be said to be of bounded variation in $(x, \xi)$ if it is of bounded variation in $\xi$ for each $x$, and in $x$ for each $\xi$, the total variation being in each case uniformly bounded on $(a, b)$. A function will be designated simply as integrable if it possesses the properties of the functions $h_{n-2}, \cdots, h_0$ below.

† The notation $A \cdot B \equiv (A \cdot B)(x, \xi)$, $A \cdot f(x)$ will be used throughout this paper to designate the "composition" of two kernels $\int_a^b A(x, s)B(s, \xi)ds$, or, respectively, the composition of a kernel and a function, $\int_a^b A(x, s)f(s)ds$. The reciprocal of a given kernel will be usually designated by the corresponding letter of the German alphabet.

‡ Hellinger-Toeplitz, *Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten*, Encyklopädie der Mathematischen Wissenschaften, $II_3$, 1927, pp. 1379–1380.

§ It is hardly necessary to specify various conditions of integrability which should be added in the statement of Lemma 1.

$$y(x) = f(x) - \mathfrak{R}\cdot f(x),$$

yields the desired result.

3. **The extension of certain results of [I].** As in [I; 15] we define

$$T_1(f; x, \xi) \equiv \rho^{n-1}\int_a^b G(x, t; \rho)p_{n0}(t)f(t, \xi)dt, \tag{3.1}$$

$$T_2(f; x, \xi) \equiv \rho^{n-1}\int_a^b G(t, x; \rho)p_{n0}(t)f(\xi, t)dt. \tag{3.2}$$

The symbol $\Lambda(u; x, \xi)$ will be used as a generic notation to designate a linear form in $n$ arguments $u_1, u_2, \cdots, u_n$,

$$\Lambda(u; x, \xi) \equiv \sum_{i=1}^{n} u_i\lambda_i(x, \xi; \rho)$$

whose coefficients $\lambda_i(x, \xi; \rho)$ are defined as sums of a finite number of products of the form

$$[\phi(x)\psi(\xi)]E(\rho),^*$$

the functions $\phi(x), \psi(\xi)$ being continuous and of bounded variation on $(a, b)$. Likewise, $\Lambda(u; x)$ will designate a linear form

$$\Lambda(u; x) \equiv \sum_{i=1}^{n} u_i\lambda_i(x; \rho)$$

whose coefficients $\lambda_i(x; \rho)$ are sums of a finite number of terms $[\phi(x)]E$, where $E$ may depend on $\rho$ and other variables, including $x$. With this notation we have [I; 15]

LEMMA 2. (i) *If $f(x, \xi)$ is any function integrable on $(a, b)$, then the integrals*

$$\rho^{-s}D_x^s T_1(f; x, \xi) \quad (s = 0, 1, \cdots, n-1); \quad T_2(f; x, \xi) \tag{3.3}$$

*tend uniformly (in $x$, $\xi$ and $\rho$) to zero as $|\rho| \to \infty$ and $\rho$ remains in $(\mathfrak{D}_\delta)$.*

(ii) *If $f(x, \xi)$ is any function of bounded variation on $(a, b)$ then*

$$\begin{aligned} |\rho^{-s}D_x^s T_1(f; x, \xi)| &< NV_f/|\rho| \qquad (s = 0, 1, \cdots, n-1); \\ |T_2(f; x, \xi)| &< NV_f/|\rho|, \end{aligned} \tag{3.4}$$

*where the constants $N$, $V_f$ have the same meaning as in* [I; 15].

(iii) *If $f(x, \xi)$ is any function whose first derivatives are of bounded variation on $(a, b)$, then, in $(\mathfrak{D}_\delta)$,*

* By $E(\rho)$, $E(\rho, x)$, $E(\rho, x, \xi)$, $\cdots$ we designate functions which remain uniformly bounded on the respective ranges of their arguments.

$$T_1(f;\, x,\, \xi) = f(x,\, \xi)/\rho + \Lambda(\tilde{\omega}_x';\, x,\, \xi)/\rho + E/\rho^2,$$
$$T_2(f;\, x,\, \xi) = f(\xi,\, x)/\rho + \Lambda(\tilde{\omega}_x'';\, x,\, \xi)/\rho + E/\rho^2, \tag{3.5}$$

*while*

$$\rho^{-s+1}D_x{}^sT_1(f;\, x,\, \xi) = \Lambda(\tilde{\omega}_x';\, x,\, \xi) + E/\rho^2 \qquad (s = 1, 2, \cdots, n-1). \tag{3.6}$$

The proof of Lemma 2 is obtained in precisely the same manner as the proof of the corresponding Lemma 1 of [I; 15]. The foot-note on page 772 of [I] should be used in deriving (3.6).*

**Remark.** Lemma 2 can be applied even when $f(x,\, \xi)$ is replaced by a function of the type $\Lambda(u;\, x,\, \xi)$ with the result

$$p_{n0}(\xi)\rho^{n-1}\int_a^b \Lambda_0(\tilde{\omega}_x';\, x,\, s)G(s,\, \xi;\, \rho)ds = \Lambda_0(\tilde{\omega}_x';\, x,\, \xi)/\rho + \sum_{i=1}^{n}\tilde{\omega}_{i\xi}''E(x;\, \rho)/\rho + \rho^{n-2}(E\cdot G)(x,\, \xi;\, \rho) + E/\rho^2, \tag{3.7}$$

the function $\Lambda_0$ being the same on both sides of (3.7).

LEMMA 3. *The following rules may be used when operating with the symbol* $\Lambda$:

(i) $\Lambda_1(\tilde{\omega};\, x,\, \xi) + \Lambda_2(\tilde{\omega};\, x,\, \xi) = \Lambda(\tilde{\omega};\, x,\, \xi)$.

(ii) *If* $f(x,\, \xi)$ *is any function of bounded variation on* $(a,\, b)$, *then*

$$\int_a^b \Lambda(\tilde{\omega}_s;\, s,\, x)f(s,\, \xi)ds = E/\rho. \tag{3.8}$$

(iii) *Under the hypothesis of* (ii)

$$\int_a^b \Lambda(\tilde{\omega}_x;\, x,\, s)f(s,\, \xi)ds = \Lambda(\tilde{\omega}_x;\, x,\, \xi), \tag{3.9}$$

(iv)
$$\int_a^b \Lambda_1(\tilde{\omega}_x;\, x,\, s)\Lambda_2(\tilde{\omega}_s;\, s,\, \xi)ds = E/\rho.$$

The rules (i) and (iii) follow immediately from the definition of the symbol $\Lambda$. Rule (ii) follows from Lemma 3 of [I; 16], and rule (iv) is readily proved by using the definition of the symbol $\Lambda$ and applying rule (ii).

4. **The formal expression for the Green's function of problem** $(\star)$. Let $G(x,\, \xi;\, \rho)$ denote, as usual, the Green's function of the associated differential boundary value problem

$$L(u) = f; \quad L_i(u) = 0 \qquad (i = 1, 2, \cdots, n).$$

* The subscripts $x$, $t$ will be used to designate the dependence of the quantities $\tilde{\omega}_t'$, $\tilde{\omega}_i''$ of [I; 14] respectively on $x$ or on $t$. If the distinction between $\tilde{\omega}'$ and $\tilde{\omega}''$ is not essential, we shall use simply $\tilde{\omega}$.

If $\rho$ is not a pole of $G(x, \xi; \rho)$ then our problem $(\star)$ is equivalent to the integral equation

$$u(x) = F(x; \rho) + H \cdot u(x; \rho)$$

where

$$F(x; \rho) \equiv G \cdot f(x), \tag{4.1}$$

$$H(x, \xi; \rho) \equiv (G \cdot h)(x, \xi; \rho). \tag{4.2}$$

If the reciprocal $\mathfrak{H}(x, \xi; \rho)$ of the kernel $H(x, \xi; \rho)$ exists, then

$$u(x) = F(x; \rho) - \mathfrak{H} \cdot F(x, \rho) = (G - \mathfrak{H} \cdot G) \cdot f(x).$$

Hence the Green's function $\Gamma(x, \xi; \rho)$ of the problem $(\star)$ is given by

$$\Gamma(x, \xi; \rho) = (G - \mathfrak{H} \cdot G)(x, \xi; \rho) = G(x, \xi; \rho) - \int_a^b \mathfrak{H}(x, s; \rho) G(s, \xi; \rho) ds. \tag{4.3}$$

Therefore the whole question is reduced to the discussion of the kernels $H(x, \xi; \rho)$ and $\mathfrak{H}(x, \xi; \rho)$. This discussion can be carried through if we restrict $\rho$ to range over the domain $(\mathfrak{D})$ [I; 23], on the basis of the known behavior of $G(x, \xi; \rho)$ on $(\mathfrak{D})$ and of the results of the preceding §§2 and 3. The restriction above concerning $\rho$ will be assumed in the sequel without being mentioned explicitly.

5. **The discussion of the equivalent integral equation.** We start with the kernel $H(x, \xi; \rho)$. On substituting expression (1.4) for $h(x, \xi; \rho)$ into (4.2) we see at once that

$$H(x, \xi; \rho) = \rho T_1(h_n; x, \xi) + T_1(h_{n-1}; x, \xi) + \sum_{s=2}^{n} \rho^{-s+1} T_1(h_{n-s}; x, \xi).$$

Hence, by Lemma 2,

$$H(x, \xi; \rho) = h_n(x, \xi) + l(\tilde{\omega}_x'; x, \xi) + E/\rho, \tag{5.1}$$

$$T_1(h_n; x, \xi) = h_n(x, \xi)/\rho + l(\tilde{\omega}_x'; x, \xi)/\rho + E/\rho^2, \tag{5.2}$$

where $l(\tilde{\omega}_x'; x, \xi)$ is an expression of type $\Lambda$.*

We proceed now to the reciprocal $\mathfrak{H}(x, \xi; \rho)$. We set in Lemma 1

$$K_1(x, \xi) = h_n(x, \xi), \quad K_2(x, \xi) = l(\tilde{\omega}_x'; x_1, \xi) + E/\rho.$$

Since the reciprocal $\mathfrak{K}_1(x, \xi) = \mathfrak{h}_n(x, \xi)$ exists by hypothesis we have only to investigate the existence of the reciprocal of the kernel

$$K_0 \equiv l(\tilde{\omega}_x'; x, \xi) - (\mathfrak{h}_n \cdot l)(x, \xi; \rho) + E/\rho = l(\tilde{\omega}_x'; x, \xi) + E/\rho.$$

* $l(\tilde{\omega}_x'; x, \xi)$ will *not* be used as a generic notation. It is completely determined by formula (5.2).

By Lemma 3 the iterated kernel

$$K_0^{(2)}(x, \xi; \rho) = E/\rho.$$

Since this can be made as small as we please by taking $|\rho|$ sufficiently large, the reciprocal $\mathfrak{K}_0''(x, \xi; \rho)$ of the kernel $K_0^{(2)}(x, \xi; \rho)$ exists in $(\mathfrak{D})$ and is itself of the type $E/\rho$. But, from

$$u = f + K_0 \cdot f = f + K_0 \cdot f + K_0^2 \cdot u,$$
$$u = f + K_0 \cdot f - \mathfrak{K}_0'' \cdot (f + K_0 \cdot f) = f + (K_0 - \mathfrak{K}_0'' - \mathfrak{K}_0'' \cdot K_0) \cdot f$$

it is readily seen that

$$\mathfrak{K}_0 = -K_0 + \mathfrak{K}_0'' + \mathfrak{K}_0'' \cdot K_0 = -l(\tilde{\omega}_x'; x, \xi) + E/\rho.$$

Substitution into (2.3) gives now

$$(5.3) \quad \mathfrak{H}(x, \xi; \rho) \equiv \mathfrak{K} = \mathfrak{h}_n(x, \xi) - l(\tilde{\omega}_x'; x, \xi) + (l \cdot \mathfrak{h}_n)(x, \xi; \rho) + E/\rho.$$

We can prove now the following

THEOREM 1. *Under the hypotheses* (**A**), (**B**), (**C**) *above, the Green's function* $\Gamma(x, \xi; \rho)$ *of the problem* (★) *is meromorphic in* $\rho$; *furthermore, in* $(\mathfrak{D})$,

$$(5.4) \quad \Gamma(x, \xi; \rho) - G(x, \xi; \rho) = -(\mathfrak{H} \cdot G)(x, \xi; \rho) = E/\rho^n.$$

The first statement of Theorem 1 is proved in precisely the same manner as the corresponding statement of Theorem 3 of [I; 21]. The second statement follows immediately from (4.3), (5.3) if we apply Lemmas 2 and 3, and Remark to Lemma 2.

COROLLARY. *Under the hypotheses of Theorem* 1, *Theorems* 5 *and* 7 *of* [I; 23, 26] *also hold.*

6. **The equiconvergence theorem.** It was stated in [I, p. 788]† that in the equiconvergence theorem the Birkhoff integral

$$\frac{1}{2\pi i} \int_{(C_R)} \rho^{n-1} d\rho \int_a^b \Gamma(x, t; \rho) p_{n0}(t) f(t) dt$$

has to be replaced in the present theory by the integral

$$(6.1) \quad I_R^*(f) \equiv \frac{1}{2\pi i} \int_{(C_R)} \rho^{n-1} d\rho \left\{ \int_a^b f(t) dt \left( \Gamma(x, t; \rho) p_{n0}(t) - \int_a^b \Gamma(x, \xi; \rho) h_n(\xi, t) p_{n0}(\xi) d\xi \right) \right\}.$$

This leads to the

† See also the paper of Lichtenstein referred to in the foot-note ‡ on page 755 of [I].

THEOREM 2. *Under the hypotheses* (A), (B), (C) *above, if* $(C_R)$ *denotes a circle in* $(\mathfrak{D})$ *about the origin, the difference*

$$I_R(f) \equiv I_R^*(f) - \frac{1}{2\pi i}\int_{(C_R)} \rho^{n-1}d\rho \int_a^b G(x, t; \rho)p_{n0}(t)f(t)dt \tag{6.2}$$

*tends to zero as* $R\to\infty$, *for any integrable function* $f(x)$, *and uniformly in* $x$ *on* $(a, b)$.

Set

$$\phi_R(x, t) \equiv \frac{1}{2\pi i}\int_{(C_R)} \rho^{n-1}d\rho\Big\{p_{n0}(t)(\Gamma(x, t; \rho) - G(x, t; \rho)) - \int_a^b \Gamma(x, \xi; \rho)h_n(\xi, t)p_{n0}(\xi)d\xi\Big\}. \tag{6.3}$$

Then

$$I_R(f) = \int_a^b \phi_R(x, t(f(t)dt, \tag{6.4}$$

and, by the classical theorem of Lebesgue,† Theorem 2 will be proved if we prove the following:

(i) The function $\phi_R(x, t)$ is uniformly bounded for $(C_R)$ in $(\mathfrak{D})$, $x$, $t$ on $(a, b)$.

(ii) As $R\to\infty$ the integral

$$\int_\alpha^\beta \phi_R(x, t)dt$$

tends to zero, uniformly in $(x, \alpha, \beta)$ on $(a, b)$.

To prove (i) we observe that the integral of the second line of (6.3) reduces to $\rho^{1-n}T_1(h_n; x, t)$ if $\Gamma(x, \xi; \rho)$ be replaced by $G(x, \xi; \rho)$. Since by Lemma 2, $T_1(h_n; x, t) = E/\rho$ and, by Theorem 1, $\Gamma - G = E/\rho^n$, we have

$$\phi_R(x, t) = \int_{(C_R)} (E/\rho)d\rho$$

which proves (i).

To prove (ii) we have, in view of (4.3) and (3.1),

$$2\pi i\phi_R(x, t) = -p_{n0}(t)\int_{(C_R)} \rho^{n-1}(\mathfrak{H}\cdot G)(x, t; \rho)d\rho - \int_{(C_R)} T_1(h_n; x, t)d\rho + \int_{(C_R)} \rho^{n-1}(\mathfrak{H}\cdot G)\cdot h_n'(x, t)d\rho, \tag{6.5}$$

† *Sur les intégrales singulières*, Annales de Toulouse, (3), vol. 1 (1909), pp. 25–117 (52, 68).

where

$$h_n'(x, \xi) = p_{n0}(x) h_n(x, \xi).$$

Now, by (5.3),

$$(\mathfrak{H} \cdot G)(x, t; \rho) = (\mathfrak{h}_n \cdot G)(x, t; \rho) - (l \cdot G)(x, t; \rho) + (l \cdot \mathfrak{h}_n \cdot G)(x, t; \rho) + (E \cdot G)(x, t; \rho)/\rho.$$

Since $(l \cdot \mathfrak{h}_n)(x, \xi; \rho)$ is of the type $\Lambda(\tilde{\omega}_x'; x, \xi)$, an easy application of Lemma 2, Remark to Lemma 2, and Lemma 3 gives

$$p_{n0}(t)\rho^n(\mathfrak{H} \cdot G)(x, t; \rho) = \mathfrak{h}_n(x, t) + \Lambda(\tilde{\omega}_t''; x, t) - l(\tilde{\omega}_x'; x, t) + (l \cdot \mathfrak{h}_n)(x, t; \rho) + \sum_{i=1}^{n} \tilde{\omega}''_{it} E(x, \rho) + \rho^{n-1}(E \cdot G)(x, t; \rho) + E/\rho.$$

Again, by (5.2),

$$\rho T_1(h_n; x, t) = h_n(x, t) + l(\tilde{\omega}_x'; x, t) + E/\rho,$$

and, finally,

$$\begin{aligned} \rho^n(\mathfrak{H} \cdot G) \cdot h_n'(x, t) &= \rho\mathfrak{H} \cdot T_1(h_n; x, t) \\ &= (\mathfrak{h}_n - l + l \cdot \mathfrak{h}_n + E/\rho) \cdot (h_n + l + E/\rho)(x, t) \\ &= (\mathfrak{h}_n \cdot h_n)(x, t) - (l \cdot h_n)(x, t; \rho) + (l \cdot \mathfrak{h}_n \cdot h_n)(x, t; \rho) + E/\rho. \end{aligned}$$

It should be observed that in these formulas we have to consider only the leading terms in the brackets of the coefficients of the linear forms $l$ and $\Lambda$, the contributions of the remaining parts being included in the expression $E/\rho$.

On collecting all these results and taking in account the reciprocity relations

$$h_n + \mathfrak{h}_n = \mathfrak{h}_n \cdot h_n = h_n \cdot \mathfrak{h}_n$$

we see at once

$$2\pi i \phi_R(x, t) = \int_{(C_R)} \left\{ \Lambda(\tilde{\omega}_t''; x, t) + \sum_{j=1}^{n} \tilde{\omega}_{jt}'' \, E(x, \rho) + \rho^{n-1}(E \cdot G)(x, t; \rho) + \frac{E}{\rho} \right\} \frac{d\rho}{\rho}.$$

It is readily seen however that

$$\int_\alpha^\beta \Lambda(\tilde{\omega}_t''; x, t) dt = \frac{E}{\rho}; \quad \int_\alpha^\beta \tilde{\omega}_{jt}'' \, dt = \frac{E}{\rho};$$

$$\int_\alpha^\beta \rho^{n-1}(E \cdot G)(x, t; \rho) dt = \frac{E}{\rho}.$$

Hence

$$\int_{\alpha}^{\beta} \phi_R(x, t)dt = \int_{(C_R)} E\rho^{-2}d\rho,$$

which yields the desired result.

The reader will find no difficulty now in extending Theorem 6 of [I; 24] as well as in applying the above results to various problems of expansion of an arbitrary function in a series of fundamental functions of the problem (★), and of the summability of such series. Our results are also readily extended to the case where the right-hand member of the integro-differential equation (1.1) has the more general form considered in [I; 19–26]. This extension (in the case where instead of a single integro-differential equation of $n$th order we deal with a system of $n$ integro-differential equations of the first order) is treated in a Thesis by F. C. Jonah, which will be published elsewhere.

## Errata corrigenda to the paper [I]

Page 769, foot-note: The integrand should be replaced by

$$\{\mathfrak{F}_1(\theta_i) - \tfrac{1}{2}(\partial\mathfrak{F}'(\theta_i)/\partial x)\}/\mathfrak{F}'(\theta_i).$$

Pages 763–764: The expressions for the pseudo-resolvent used in the text are correct only in the case of real kernels. Since it is essential here to consider complex-valued kernels as well, the following modifications should be made: $\omega_k'$, $\omega_k$ of the last formula on page 763, and in the formulas on lines 3, 11, 16, page 764, should be replaced by their conjugates, $\bar{\omega}_k'$, $\bar{\omega}_k$. The same remark concerns $\omega_s$ on 8th line.

Pages 774–775: The right-hand member of (29) should be replaced by

$$\sum_j [\omega_{ik0j}(x, t)]E_{ij}(\rho),$$

and $\omega_{ik0}$, $E_{ik}$ in the next two lines should be replaced by $\omega_{ik0j}$, $E_{ij}$ respectively. An analogous correction should be made in the formula (34). In (30) $\bar{\omega}_i$ should be replaced by $\bar{\omega}_i'$.

Pages 785–786: It is not always possible to draw the family of contours $(C_R)$ in $(\mathfrak{T})$ in such a way that one and only one pole of the Green's function be enclosed between any two consecutive contours. Hence the terms in the series expansion of the Corollary of 25 should be grouped, each group containing the terms that correspond to sets of poles between the consecutive contours $(C_R)$. This situation, which was stated correctly in the author's paper referred to in the foot-note § on page 755, is overlooked in [I] and in the author's paper referred to by "D" in [I]. The same remark, apparently, should be made concerning all the publications on the subject by various other authors.

Brown University,
Providence, R. I.

# ANALYTIC FUNCTIONS IN THE IRREGULAR FIELD OF ALL NUMERICAL FUNCTIONS*

BY

E. T. BELL

1. **Introduction.** To distinguish them from the elements with which we shall be chiefly concerned, we shall call real or complex numbers *scalars*. As it is usually immaterial in the sequel whether the scalars are real or imaginary, either may be understood unless the contrary is expressed. Scalars will be denoted by small Latin letters, $a, b, \cdots, x, y, \cdots$.

For the moment only, let $\alpha, \beta, \cdots, \xi, \eta, \omega$ denote elements of an abstract field $F$ whose zero, unit elements are $\omega, \eta$ respectively, and in which sums, products, quotients, differences are written as usual, $\alpha+\beta$, $\alpha\beta$, $\alpha/\beta$, $\alpha-\beta$. The properties of products such as $a\alpha$ are assumed in the customary form.† To the classical postulates and definitions of $F$, it is convenient for our purpose to add the following definition: An element $\xi$ of $F$ is called *regular* or *irregular* according as there exists or does not exist a uniquely defined element $\xi'$ in $F$, called the *reciprocal* of $\xi$, such that $\xi\xi'=\eta$. The unique irregular element in $F$ is $\omega$.

Suppose now that elements $\alpha, \beta, \cdots$ and the four fundamental operations upon them (addition, subtraction, multiplication, division) are defined so that the resulting system $IF$ differs from $F$ only in the existence in $IF$ of an infinity of irregular elements. Then $IF$ is called an *irregular field*. The postulates for $IF$ were precisely stated in a former paper,‡ and numerous instances were given there and elsewhere.§ The foregoing description of $IF$ is sufficient for the present, as we shall discuss an instance of $IF$ in some detail. In the papers cited the algebra (theory of finite processes) in $IF$ was fully treated; here, we consider the analysis (theory of essentially infinite processes) so far as it relates to functions in $IF$ having power series expansions in $IF$, for the particular instance mentioned. Elements of the $IF$ considered will be denoted by small Greek letters, $\alpha, \beta, \cdots, \omega, \eta, \xi, \cdots$.

Before proceeding, it may be of interest to state an example of the similarities and differences between analytic functions in a scalar field and

* Presented to the Society, April 5, 1930; received by the editors in February, 1930.

† See, for example, L. E. Dickson, *Algebren und ihre Zahlentheorie*, 1927, pp. 23–24.

‡ E. T. Bell, Annals of Mathematics, vol. 27 (1926), p. 511.

§ E. T. Bell, *Algebraic Arithmetic*, 1927.

their correspondents in $IF$. It was shown by Tannery* that if $S(x)$, $C(x)$ are uniquely defined for all values of $x$, and have the addition theorems

$$\begin{aligned} S(x+y) &= S(x)C(y) + C(x)S(y), \\ C(x+y) &= C(x)C(y) - S(x)S(y), \end{aligned} \tag{1.1}$$

and if there exists a single value of $x$ for which $C(x)$, $S(x)$ are continuous, then they are continuous for all $x$, and, the trivial case where both vanish identically being excluded,

$$S(x) = e^{ax}\sin bx, \qquad C(x) = e^{ax}\cos bx, \tag{1.2}$$

where $a$, $b$ are constants. If in addition

$$S^2(x) + C^2(x) = 1, \tag{1.3}$$

then $S(x)=\sin bx$, $C(x)=\cos bx$. Further instances of completely solved systems of functional equations will be found in Osgood (loc. cit.) and in Van Vleck and H'Doubler.† Remarks similar to the following, for the above set, apply to all.

We note first that the values of $S(x)$, $C(x)$ are scalars, as also is $x$. We shall require our functions in $IF$, that is their "values," to be in the same $IF$ as their arguments. Second, if $x$ is not scalar, continuity is not significant, and hence must be dropped in $IF$. Third, the right-hand member of (1.3) is the unit scalar; in $IF$ it must be replaced by $\eta$, the unit element in $IF$, which is such that, if $\xi$ is any element of $IF$, $\eta\xi=\xi$. Fourth, the (scalar) solution of (1.1), (1.3) having been shown to exist under the stated hypotheses, the power series for the sine and cosine follow in the usual way, and are absolutely convergent for all finite values of the argument.

In $IF$ the situation corresponding to (1.1), (1.3) is

$$\begin{aligned} S(\alpha+\beta) &= S(\alpha)C(\beta) + C(\alpha)S(\beta), \\ C(\alpha+\beta) &= C(\alpha)C(\beta) - S(\alpha)S(\beta), \end{aligned} \tag{1.11}$$

where $S(\xi)$, $C(\xi)$ are to be uniquely defined for every $\xi$ in $IF$; and

$$S^2(\xi) + C^2(\xi) = \eta. \tag{1.31}$$

The conclusion is that

$$S(\xi) = \sum_{r=0}^{\infty} \frac{(-1)^r\xi^{2r+1}}{(2r+1)!}, \qquad C(\xi) = \sum_{r=0}^{\infty} \frac{(-1)^r\xi^{2r}}{(2r)!}; \tag{1.4}$$

* See, for example, W. F. Osgood, *Lehrbuch der Funktionentheorie*, 1926, pp. 582–586.

† E. B. Van Vleck and F. H' Doubler, these Transactions, vol. 17 (1916), pp. 9–49.

$$S(-\xi) = -S(\xi), \quad C(-\xi) = C(\xi); \qquad S(\omega) = \omega, \quad C(\omega) = \eta, \tag{1.5}$$

where $\omega$ is the (unique) zero element in $IF$, defined by $\omega\xi = \omega$ for all $\xi$ in $IF$ ($\omega$ also is irregular).

Further, from the definition of derivatives in $IF$ with respect to $\xi$, indicated by $D_\xi$, it follows that $D_\xi S(\xi) = C(\xi)$, $D_\xi C(\xi) = -S(\xi)$. The next is in striking contrast to the scalars $S(x)$, $C(x)$. In the instance of $IF$ considered, there is uniquely determined a constant finite scalar, say $p$, when $\xi$ is a designated element of $IF$, such that $S(\xi)$, $C(\xi)$ as in (1.4) are reducible to the forms

$$S(\xi) = C'(\xi) \sin p + S'(\xi) \cos p,$$
$$C(\xi) = C'(\xi) \cos p - S'(\xi) \sin p,$$

where $C'(\xi)$, $S'(\xi)$ are certain finite linear homogeneous functions of certain elements of $IF$ which are uniquely determined when $\xi$ is given, and which further satisfy the same functional equations (1.11), (1.31), (1.5) as $C(\xi)$, $S(\xi)$. The sense in which infinite processes exist, or are convergent, in $IF$ will appear as we proceed.

The similarities and differences illustrated above arise from two sources: the existence in $IF$ of an infinity of irregular elements, and the specific interpretations assigned to the four fundamental operations.

2. **The $IF$ of all numerical functions.** We shall recall from previous papers a minimum of definitions and theorems to make the present self-contained.

If for all finite integral values $>0$ of the scalar variable $x$ the scalar function $\xi(x)$ is finite and uniform, $\xi(x)$ is called a *numerical function of* $x$, and is said to be *regular* or *irregular* according as $\xi(1) \neq 0$ or $\xi(1) = 0$. The *unit*, *zero* numerical functions, $\eta(x)$, $\omega(x)$ are defined by

$$\eta(1) = 1, \quad \eta(n+1) = 0, \quad \omega(n) = 0 \qquad (n = 1, 2, \cdots).$$

The numerical functions $\alpha(x)$, $\beta(x)$ are defined to be *equal* when and only when $\alpha(n) = \beta(n)$ for all integers $n > 0$. If $\alpha(x)$, $\beta(x)$ are equal, the equality is written $\alpha = \beta$. Hence the following are equivalent statements:

$$\alpha = \beta, \quad \alpha(n) = \beta(n) \qquad (n = 1, 2, \cdots). \tag{2.1}$$

As such symbols $\alpha, \beta, \cdots$ occur only in relations of equality as just defined, it is unnecessary to define them separately. The special irregular field $IF$ to be discussed has as its *elements* the set of all $\alpha, \beta, \cdots, \eta, \omega, \xi, \cdots$, where $\alpha(x), \beta(x), \cdots, \xi(x), \cdots$ is the set of all numerical functions; $\eta(x)$, $\omega(x)$ are as above defined, and $\xi$ is *regular* or *irregular* according as $\xi(x)$ is regular or

irregular. To avoid repeating $n=1, 2, \cdots$, we make the convention that any relation involving $n$ is to hold for all integral values $>0$ of $n$. The fundamental operations in $IF$ are defined by the following pairs of statements, postulated to be equivalent:

$$(2.2) \qquad \phi = c\xi, \qquad \phi(n) = c\xi(n);$$

$$(2.3) \qquad \phi = \alpha + \beta, \qquad \phi(n) = \alpha(n) + \beta(n);$$

$$(2.4) \qquad \phi = \alpha\beta, \qquad \phi(n) = \sum\alpha(d)\beta(t),$$

the summation referring to all distinct matrices $(d, t)$ of integers $d>0$, $t>0$ such that $dt=n$.

It was shown elsewhere* that if and only if $\xi$ is regular, there exists in $IF$ a uniquely determined element $\xi'$, called the *reciprocal* of $\xi$, such that $\xi\xi'=\eta$, and an explicit form for $\xi'(n)$ in terms of $\xi(d)$, where $d$ runs through the positive integral divisors of $n$, was given. Here it is sufficient to know that $\xi'$ exists when and only when $\xi$ is regular. Writing $\xi'=\eta/\xi=\xi^{-1}$ ($\xi$ regular), we thus define division in $IF$.

From (2.2), (2.3), $a\alpha+b\beta+\cdots+c\gamma$ is defined, and from (2.4) the definition of $\xi_1\xi_2\cdots\xi_r$ follows:

$$\phi = \xi_1\xi_2\cdots\xi_r, \qquad \phi(n) = \sum\xi_1(n_1)\xi_2(n_2)\cdots\xi_r(n_r),$$

the summation referring to all matrices $(n_1, n_2, \cdots, n_r)$ of integers $n_j>0$ $(j=1, \cdots, r)$ such that $n=n_1n_2\cdots n_r$. If $\xi_1=\xi_2=\cdots=\xi_r$, we write $\phi=\xi^r$. If $\xi$ is regular, and $\xi'$ is the reciprocal of $\xi$, $\xi^{-r}$ is defined by $\xi^{-r}=\xi'^r$. If $\alpha$ is any element ($\omega$ not excluded) of $IF$, by definition $\alpha^0=\eta$. If $\xi$ is any element of $IF$, $\eta\xi=\xi$, $\omega\xi=\omega$; the unit element $\eta$ is unique, as also is the zero element $\omega$. The properties of scalar multiplication defined in (2.2), addition in (2.3) and multiplication in (2.4) are abstractly identical with those of the similarly named operations in a field, and hence with respect to these operations the elements of $IF$ form a ring, say $IR$. Adjoining to $IR$ division in $IF$ as above defined, we see that $IF$ is abstractly identical with common algebra, except that in $IF$ division is impermissible by each of an infinity of elements of $IF$ (the irregular elements), instead of merely by the unique zero. Details, if desired, will be found in the papers cited; the foregoing is sufficient for what follows.

3. **Components.** Throughout this section, $\xi$ is an arbitrary element of $IF$, $r$ is an arbitrary integer $\geqq 0$, and $p=\xi(1)$. We shall define elements $\xi_a(a=0, \cdots, r)$ in $IF$, the *components* of $\xi^r$, such that

* E. T. Bell, Tôhoku Mathematical Journal, 1921.

$$\xi^r = r! \sum_{a=0}^{r} \frac{p^{r-a}}{(r-a)!} \xi_a. \tag{3.1}$$

It is easily shown that if the $c_a$ are scalars, and the $\xi_a'$ elements of $IF$, a resolution of the form

$$\xi^r = \sum_{a=0}^{r} c_a \xi_a'$$

is, if it exists, unique. Thus (3.1), when the $\xi_a$ are determined, will be the unique resolution of $\xi^r$ into its components in $IF$. If $s$ is an integer $<0$, and $\xi$ is irregular, $\xi^s$ does not exist; if $\xi$ is regular, $\xi^s = (\xi')^{-s}$, $-s>0$, $\xi' =$ the reciprocal (in $IF$) of $\xi$, so that this case is included in (3.1). Hence (3.1) is the general case. The resolution of powers into components enables us to pass, as will be seen later, directly from finite to infinite processes in $IF$. We proceed to find the components $\xi_a$.

Let $n = q_1^{t_1} \cdots q_j^{t_j} (t_1 > 0, \cdots, t_j > 0)$ be the resolution of $n>1$ into a product of powers of distinct primes $q_1, \cdots, q_t$ (1 is not a prime); write $\pi(1) = 0$, $\pi(n) = t_1 + \cdots + t_j$. Then no resolution of the integer $m>0$ into more than $\pi(m)$ factors $>1$ is possible.

By definition, we take

$$\xi_a(n) = 0 \qquad (a > \pi(n),\ n = 1, 2, \cdots); \tag{3.2}$$

$$\xi_0 = \xi^0 = \eta; \quad \xi_b = \omega \qquad (b < 0). \tag{3.3}$$

Hence we need consider only $\xi_a(n)$, $n>1$, $0<a \leqq \pi(n)$.

Let $n$ be an integer $>1$; let $a_1, a_2, \cdots, a_s$ be $s$ integers $>0$, and $P_1, P_2, \cdots, P_s$ be $s$ distinct integers $>1$ which, without loss of generality, may be assumed in ascending order, $P_1 < P_2 < \cdots < P_s$, such that

$$n = P_1^{a_1} P_2^{a_2} \cdots P_s^{a_s}. \tag{3.4}$$

Write $a = a_1 + a_2 + \cdots + a_s$. We shall call (3.4) a *decomposition* of $n(>1)$ of *degree* $a$.

Consider $\xi^r(n) (n>1)$. If $b>r$, no decomposition of $n$ of degree $b$ contributes to $\xi^r(n)$. By the definition in §2 of multiplication in $IF$,

$$\xi^r(n) = \sum \xi(n_1)\xi(n_2) \cdots \xi(n_r),$$

$\sum$ having the signification there stated. Hence, if we define $\xi_a(n)$ to be the sum

$$\sum \frac{(\xi(P_1))^{a_1}(\xi(P_2))^{a_2} \cdots (\xi(P_s))^{a_s}}{a_1! a_2! \cdots a_s!} \tag{3.5}$$

extended over all decompositions of $n$ of degree $a$ when $a \leqq r$, the total contribution of all the decompositions of degree $a$ of $n$ to $\xi^r(n)$ is

$$\frac{r!p^{r-a}}{(r-a)!}\xi_a(n), \tag{3.6}$$

since, for example, the particular decomposition (3.4) must be written

$$n = 1^{r-a}P_1^{a_1}P_2^{a_2}\cdots P_s^{a_s}$$

in order to contribute to $\xi^r(n)$.

Finally, then, with $\xi_a(n)(a \geqq 0,\ n>0)$ as defined in (3.2), (3.3), (3.5), we see by these and (3.6) that the required resolution of $\xi^r(r \geqq 0)$ into its components $\xi_a$ is (3.1).

In discussing infinite processes in $IF$, all powers are resolved into their components. From the definitions, all products of components are products in $IF$. The last remark may be emphasized, as the final theorems derived from power series in $IF$ are of precisely the same forms as their correspondents in a scalar field.

In (3.1) we have resolved a power into its components. There is a reciprocal resolution of $\xi_r$ into a polynomial in $\xi$ of degree $r$. Writing (3.1) for the moment in the form

$$\xi^r = \sum_{a=0}^{r} M_a(r)p^{r-a}\xi_a, \tag{3.7}$$

we observe that the $M_a(r)$ are rational numbers depending only upon $a$, $r$ and that the sequence $\xi_0, \xi_1, \cdots$ is uniquely determined in $IF$ when $\xi$ is given. In (3.7) replace $r$ by $0, 1, \cdots, r$, and solve the resulting system of equations (of determinant 1) for the $\xi_a$. Then

$$\xi_r = \sum_{a=0}^{r} N_a(r)\xi^a, \tag{3.8}$$

where the $N_a(r)$ are rational numbers depending only upon $a$, $r$, and the representation (3.8) is unique. The explicit forms of the $N_a(r)$ will not be required.

For completeness, we state the components of sums and products. Let $\alpha, \beta$ be any elements of $IF$. Write $\phi = \alpha+\beta$, $\psi = \alpha\beta$. We seek $\phi_r$, $\psi_r$, where $r$ is an integer $\geqq 0$. The results are

$$(\alpha+\beta)_r = \sum_{a=0}^{r} \alpha_a\beta_{r-a}, \tag{3.9}$$

as may be proved directly from the definition of $\xi_a(n)$, or immediately from the exponential theorem as in §6;

$$\psi_r = (\alpha\beta)_r = \sum_{a=0}^{r} N_a(r)\alpha^a\beta^a;$$

and, therefore,

$$(3.10) \qquad (\alpha\beta)_r = \sum_{a=0}^{r}\sum_{b=0}^{a}\sum_{c=0}^{a} N_a(r)M_b(a)M_c(a)\alpha_b\beta_c.$$

The following is a useful relation. Let $x$ be an arbitrary scalar, $\xi$ an arbitrary element of $IF$. Then

$$(3.11) \qquad (x\xi)_r = x^r\xi_r.$$

4. **Derivatives in $IF$.** The following definitions enable us to attain complete isomorphism between derivatives of functions in $IF$ and the like for scalars. The *derivatives* $D_\xi\xi^a$, $D_\xi\xi_a$ of $\xi^a$, $\xi_a$, where $a$ is an integer $\gtreqless 0$, and $\xi$ is regular, are the respective elements $a\xi^{a-1}$, $\xi_{a-1}$ of $IF$,—that is,

$$(4.1) \qquad D_\xi\xi^a = a\xi^{a-1},\ D_\xi\xi_a = \xi_{a-1}\ (a \gtreqless 0).$$

If $\xi$ is irregular, (4.1) by definition is restricted to the case $a>0$. From (4.1) there follow easily all the formal properties of derivatives of sums, products, quotients, etc., in $IF$, in complete isomorphism with the like for scalars.

5. **Analytic functions in $IF$.** Let $\xi$ be an arbitrary element of $IF$, and let $R$ be the region of convergence of $f(x)$, the scalar function defined in $R$ by

$$f(x) = \sum_{r=0}^{\infty} c_r x^r.$$

Then we define $\sum_{r=0}^{\infty} c_r\xi^r$ to be the *analytic function $f(\xi)$ of $\xi$ in $IF$ if and only if $\xi(1)$ lies within $R$.*

Write $f_0(x)=f(x)$, $f_a(x)=d^af(x)/dx^a (a>0)$, $p=\xi(1)$. Then, by the above definition, if $f(\xi)$ is an analytic function of $\xi$ in $IF$; the series

$$\sum_{j=0}^{\infty} c_{a+j}\frac{(a+j)!}{j!}p^j$$

is absolutely convergent, and its sum is $f_a(p)(a\geqq 0)$.

Let $n$ be an arbitrary integer $>0$. Then (see §3), the resolution of $\xi^r(n)$ contains at most $\pi(n)$ components $\xi_a(n)$. Denote $f(\xi)$ by $\Xi$. Then, if $\Xi$ is analytic in $IF$,

$$\begin{aligned}\Xi(n) &= \sum_{r=0}^{\infty} c_r\xi^r(n) \\ &= \sum_{r=0}^{\infty} r!c_r \sum_{a=0}^{r} \frac{p^{r-a}}{(r-a)!}\xi_a(n);\end{aligned}$$

and the last is the sum of not more than $\pi(n)$ absolutely convergent series,

the $\xi_a(n)$ being numerical constants with respect to the summations. Performing the summations, we have

$$\Xi(n) = \sum_{a=0}^{\tau(n)} f_a(p)\xi_a(n) \qquad (n = 1, 2, \cdots);$$

or, what is the same, by the definition of equality in $IF$,

$$f(\xi) = \sum_{a=0} f_a(p)\xi_a,$$

in which the upper limit of the summation may if desired be taken as $\infty$.

To summarize: *If $\xi$ is any element of $IF$, and if $\xi(1)$ lies within the region of convergence of the scalar function $f(x)$, where*

$$f(x) = c_0x^0 + c_1x + c_2x^2 + \cdots,$$

*then, by definition, $f(\xi)$, where*

$$f(\xi) = c_0\eta + c_1\xi + c_2\xi^2 + \cdots,$$

*is an analytic function of $\xi$ in $IF$, and it follows that*

$$f(\xi) = \sum_{a=0} f_a(p)\,\xi_a \qquad (p = \xi(1)), \tag{5.1}$$

*where $f_a(p)$ is the scalar $[d^af(x)/dx^a]_{x=p}$.*

Denote by $f_r(x)$ $(r \geqq 0)$ the $r$th derivative of $f(x)$ with respect to $x$. Then, if $x_0$ lies within $R$, the series

$$\sum_{j=0}^{\infty} c_{r+j}\frac{(r+j)!}{j!}x_0^j$$

is absolutely convergent and its sum is $f_r(x_0)$. Hence $f_r(\xi)$, where

$$f_r(\xi) = \sum_{j=0}^{\infty} c_{r+j}\frac{(r+j)!}{j!}\xi^j,$$

is an analytic function in $IF$ of $\xi$, and, from what precedes and §4, we have

$$f_r(\xi) = D_\xi^r f(\xi) = \sum_{a=0} f_a(p)\xi_{a-r}, \tag{5.2}$$

where $D_\xi^r$ indicates $r$ successive derivations with respect to $\xi$.

From (5.1), (5.2), we have the MacLaurin expansion of any analytic function $f(\xi)$ in $IF$,

$$f(\xi) = \sum_{a=0}^{\infty} \frac{\xi^a}{a!} f_a(\omega). \tag{5.3}$$

Similarly, if $\xi$, $\alpha$ are any elements of $IF$ such that $\xi(1)-\alpha(1)$ lies within the region of convergence of $f(x)$, we have the analogue in $IF$ of Taylor's theorem for scalars,

$$f(\xi) = \sum_{a=0}^{\infty} \frac{(\xi - \alpha)^a}{a!} f_a(\alpha). \tag{5.4}$$

From the definition of analytic functions in $IF$, it is in short clear that the *analysis of scalar analytic functions for values of their arguments within the complete intersection of the respective regions of convergence of the scalar functions goes over unchanged to IF by a mere reinterpretation of the scalar variables as elements of IF and of the four fundamental operations upon scalars as the corresponding operations in IF.* (In division it is necessary to avoid irregular divisors.)

As a detail in the transposition from scalars to their unique correspondents in $IF$, we may repeat a remark in the introduction: The scalar unit 1 corresponds to $\eta$. For example, $\cos^2\xi+\sin^2\xi=\eta$ is true; $\cos^2\xi+\sin^2\xi=1$ ismeaningless, since $\cos^2\xi$, $\sin^2\xi$ are in $IF$, and 1 is not. Similarly for

$$\sin\xi + \sin(-\xi) = \omega, \qquad \sin\xi + \sin(-\xi) = 0,$$

the first of which is true and the second meaningless.

6. **The elementary functions in $IF$.** In illustration of §5, we shall next briefly consider the exponential, circular and hyperbolic functions in $IF$. Since $p(=\xi(1))$ lies within the region of convergence of each of $\exp x$, $\sin x$, $\cos x$, $\operatorname{sh} x$, $\operatorname{ch} x$ (because $\xi(1)$ is, by the definition of numerical functions, finite), each of $\exp\xi, \cdots, \operatorname{ch}\xi$, where $\xi$ is an arbitrary element of $IF$, is an analytic function of $\xi$ in $IF$. We have

$$\exp\xi = \sum_{r=0}^{\infty} \frac{\xi^r}{r!};$$

$$\sin\xi = \sum_{r=0}^{\infty} \frac{(-1)^r\xi^{2r+1}}{(2r+1)!}, \qquad \cos\xi = \sum_{r=0}^{\infty} \frac{(-1)^r\xi^{2r}}{(2r)!}; \tag{6.1}$$

$$\operatorname{sh}\xi = \sum_{r=0}^{\infty} \frac{\xi^{2r+1}}{(2r+1)!}, \qquad \operatorname{ch}\xi = \sum_{r=0}^{\infty} \frac{\xi^{2r}}{(2r)!}.$$

Hence, if we define $E(\xi), \cdots, S(\xi)$ by

$$E(\xi) = \sum_{a=0}\xi_a, \quad A(\xi) = \sum_{a=0}\xi_{2a}, \quad B(\xi) = \sum_{a=0}\xi_{2a+1},$$

$$C(\xi) = \sum_{a=0}(\xi_{4a} - \xi_{4a+2}), \quad S(\xi) = \sum_{a=0}(\xi_{4a+1} - \xi_{4a+3}),$$

an immediate application of (5.1) gives the corresponding reduced forms,

$$
\begin{aligned}
\exp \xi &= E(\xi) \exp p; \\
\sin \xi &= C(\xi) \sin p + S(\xi) \cos p, \\
\cos \xi &= C(\xi) \cos p - S(\xi) \sin p; \\
\operatorname{sh} \xi &= A(\xi) \operatorname{ch} p + B(\xi) \operatorname{sh} p, \\
\operatorname{ch} \xi &= A(\xi) \operatorname{sh} p + B(\xi) \operatorname{ch} p.
\end{aligned} \tag{6.2}
$$

From either (6.1) or (6.2) follow all the usual properties, except those concerning continuity and scalar periodicity, which are meaningless here, of scalar exponential, circular and hyperbolic functions for the corresponding analytic functions in $IF$ in exact isomorphism. Suppose these are developed. Then we may, at once, write down a second isomorphism, as follows:

| $\exp x$ | $\sin x$ | $\cos x$ | $\operatorname{sh} x$ | $\operatorname{ch} x$ |
|---|---|---|---|---|
| $\exp \xi$ | $\sin \xi$ | $\cos \xi$ | $\operatorname{sh} \xi$ | $\operatorname{ch} \xi$ |
| $E(\xi)$ | $S(\xi)$ | $C(\xi)$ | $B(\xi)$ | $A(\xi)$ |

corresponding functions being in the same column of the table, the complete isomorphism being such that, if any given relation holds between the members of a particular row of the table, say the first, as it is the most familiar, then the relation holds also when the members of that row are replaced by their correspondents in each of the remaining rows. The isomorphism includes derivatives, $D_\xi$ being the correspondent of $d/dx$.

In the preceding theory, we started from the fundamental theorem of arithmetic (unique decomposition into primes). There is a generalization to functions of the elements in any commutative semi-group, as may be seen from the algebra developed by the author.* The ultimate interpretation of properties of analytic functions is, as I will show elsewhere, a theory of compound partitions of matrices of rational integers.

* E. T. Bell, *Algebraic Arithmetic*, chapter IV.

California Institute of Technology,
Pasadena, Calif.

# THE THEOREM OF DEDEKIND IN THE IDEAL THEORY OF ZOLOTAREV*

BY

H. T. ENGSTROM†

## I. INTRODUCTION

Let $K(\theta)$ be the algebraic field generated by a root $\theta$ of the irreducible equation

$$f(x) = x^n + a_1x^{n-1} + \cdots + a_n = 0$$

and let $p$ be a rational prime. The theorem of Dedekind on the connection between ideals and higher congruences which is fundamental in the theory of algebraic numbers is the following:

THEOREM 1. *If $p$ is not a divisor of the index of $\theta$, and*

$$f(x) \equiv \Phi_1(x)^{e_1} \cdot \Phi_2(x)^{e_2} \cdots \Phi_r(x)^{e_r} \pmod{p}$$

*is the decomposition of $f(x)$ in prime functions* (mod $p$) *then*

$$p = \mathfrak{p}_1^{e_1} \cdot \mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_r^{e_r} .$$

*is the prime ideal decomposition of $p$ in $K(\theta)$, and $N\mathfrak{p}_i = p^{f_i}$ where $f_i$ is the degree of $\Phi_i(x)$. Furthermore*

$$\mathfrak{p}_i = (p, \Phi_i(\theta)).$$

This theorem was first used by Zolotarev‡ in 1874 as a definition of ideals. This definition is not general on account of the exceptional character of the index divisors, for the ideal decomposition on this definition depends on the particular equation used to define the field. Dedekind first attempted to establish a definition of ideals in terms of the equation defining the field, but for a general theory he was forced to his abstract definition which is now classical.§ He proved Theorem 1 on this basis in 1878.¶ Zolotarev also gave

* Presented to the Society, November 29, 1930; received by the editors in January, 1930.

† National Research Fellow, California Institute of Technology.

‡ G. Zolotarev, *Théorie des nombres entiers complexes, avec une application au calcul intégral*, Bulletin de l'Académie des Sciences, St. Petersburg, 1874. Abstracted in Fortschritte der Mathematik, vol. 6 (1874), p. 177.

§ Dirichlet, *Vorlesungen über Zahlentheorie*, edited by Dedekind, 4th edition, §72–85 and §141.

¶ R. Dedekind, *Über den Zusammenhang zwischen der Theorie der Ideale und der Theorie der höheren Kongruenzen*, Göttinger Abhandlungen, 1878, p. 15.

up his attempt to define ideals directly in terms of the equation defining the field and in 1880* he succeeded in giving a general definition. On the basis of this definition Theorem 1 is no longer obvious. In the following paper I give a simple proof of the theorem by first developing the theory of congruences for ideal moduli in terms of the Zolotarev ideals.

The problem of determining the prime ideal decomposition of a rational prime $p$ in the general case was solved by Kronecker's theory of forms.† Hensel‡ in 1894 showed that in this theory there exists an analogy to the Theorem of Dedekind which is valid for all primes. The calculation of the prime ideal decomposition by this method, however, is based on an integral base of the field. In the Zolotarev theory the ideal decomposition may also be calculated for any $p$ from an integral base of the field.§ The problem of determining the ideal decomposition of $p$ directly from the equation defining the field is solved by the Theorem of Ore.‖ In a later paper I shall establish these results directly from the Zolotarev definition.

## II. Fundamental definitions and theorems on Zolotarev ideals

Greek letters are used to represent the integers of an algebraic field $K$ of $n$th degree. Italic letters represent rational integers and $p$ a rational prime.

Definition 1. *$\omega_1$ is said to divide $\omega_2$* (mod $p$) *if there exists a $c$, prime to $p$, such that $c\cdot\omega_2/\omega_1$ is an integer of $K$.*

The notation $\omega_1|\omega_2$ will be used to denote the divisibility of $\omega_2$ by $\omega_1$ (mod $p$).

Definition 2. *If $\epsilon|1$ then $\epsilon$ is a unit* (mod $p$).

Definition 3. *If $\omega_1|\omega_2$ and $\omega_2|\omega_1$ then $\omega_1$ and $\omega_2$ are called associates* (mod $p$).

In questions of divisibility (mod $p$) associates are not regarded as distinct.

---

* G. Zolotarev, *Sur la théorie des nombres complexes*, Journal de Mathématiques, (3), vol. 6 (1880), pp. 51–84, 129–166. For a modern account of the theory see N. Tchebotarev, *The foundations of the ideal theory of Zolotarev*, American Mathematical Monthly, vol. 37 (1930), pp. 117–128.

† L. Kronecker, *Grundzüge einer arithmetischen Theorie der algebraischen Grössen*, Journal für Mathematik, vol. 92 (1882), pp. 1–122.

‡ K. Hensel, *Untersuchungen der Fundamentalgleichung einer Gattung für reelle Primzahl als Modul und Bestimmung der Theiler ihrer Diskriminante*, Journal für Mathematik, vol. 113 (1894), pp. 61–83.

§ Cf. N. Tchebotarev, loc. cit.

‖ Ö. Ore, *Über den Zusammenhang zwischen den definierenden Gleichungen und der Idealtheorie in algebraischen Körpern*, Mathematische Annalen, vol. 98 (1927), Theorem 9, p. 585.

DEFINITION 4. *If $\pi$ is not a unit* (mod $p$) *and $\pi \mid \omega_1\omega_2$ only when $\pi \mid \omega_1$ or $\pi \mid \omega_2$ then $\pi$ is said to be a prime* (mod $p$).

THEOREM 2. *Each integer in $K$ has a unique decomposition in primes* (mod $p$).

THEOREM 3. *There exists a complete residue system* (mod $p$) *in $K$,*

$$\alpha_0 = p, \alpha_1, \quad \alpha_2, \cdots, \alpha_\sigma \quad (\sigma = p^n - 1), \tag{1}$$

*such that $\alpha_i \mid p$, $i = 1, 2, \cdots, \sigma$.*

THEOREM 4. *All prime divisors* (mod $p$) *of $p$ are contained in the set* (1).

DEFINITION 5. *For each rational prime $p$, to each prime divisor $\pi$*(mod $p$) *we associate a symbol $\mathfrak{P}$, called a prime ideal divisor, and say that an integer $\omega$ contains the ideal divisor $\mathfrak{P}^k$ if $\omega$ is divisible* (mod $p$) *by $\pi^k$.*

Corresponding to each rational prime dividing $N(\omega)$ as a modulus, $\omega$ will contain certain ideal divisors. We write $\omega$ symbolically as the product of all ideal divisors contained in $\omega$, i.e.,

$$\omega = \mathfrak{P}_1^{e_1}\mathfrak{P}_2^{e_2}\cdots\mathfrak{P}_r^{e_r}.$$

THEOREM 5. *Each integer in $K$ has a unique decomposition in prime ideal divisors.*

An integer $\omega$ of the field is determined to an absolute algebraic unit by its divisors for the rational primes which divide its norm as moduli. Theorem 5 is a symbolic statement that a representation of an integer $\omega$ by its divisors for the rational primes dividing $N(\omega)$ as moduli is unique (cf. Tchebotarev, loc. cit. p. 125).

DEFINITION 6. *If $\mathfrak{P}$ is the prime ideal divisor associated with the prime divisor $\pi$* (mod $p$), *the power of $p$ dividing $N(\pi)$ is called the norm, $N(\mathfrak{P})$, of $\mathfrak{P}$.*

DEFINITION 7. *The norm of a product of prime ideal divisors is the product of the norms of the factors.*

## III. CONGRUENCES FOR IDEAL MODULI

Let $\omega_1, \omega_2, \cdots, \omega_n$ be an integral base of $K$ and $\pi$ an arbitrary integer of $K$. Then

$$\pi\cdot\omega_j = a_{1j}\omega_1 + a_{2j}\omega_2 + \cdots + a_{nj}\omega_n \qquad (j = 1, 2, \cdots, n), \tag{2}$$

where $N(\pi) = \pm |a_{ij}|$. We determine a normal base for multiples of $\pi$ as follows. Of all multiples of $\pi$ in the form

$$c_1\omega_1 + c_2\omega_2 + \cdots + c_i\omega_i, \quad c_i \not\equiv 0, \tag{3}$$

let $\Omega_i$ be that for which $|c_i|$ is least. Since $N(\pi)\omega_i$ is of the form (3) such integers exist. We obtain the set

$$\begin{aligned} \Omega_1 &= b_{11}\omega_1, \\ \Omega_2 &= b_{21}\omega_1 + b_{22}\omega_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \Omega_n &= b_{n1}\omega_1 + b_{n2}\omega_2 + \cdots + b_{nn}\omega_n. \end{aligned} \tag{4}$$

It is readily shown that every multiple of $\pi$ in $K$ is a linear combination of $\Omega_1, \Omega_2, \cdots, \Omega_n$ with rational coefficients and conversely. Furthermore, if a multiple of $\pi$ has the form (3) then $c_i \equiv 0 \pmod{b_{ii}}$. Since the integers (2) also form a base for multiples of $\pi$, the determinants $|a_{ij}|$ and $|b_{ij}|$ are each divisible by the other, i.e., $N(\pi) = \pm b_{11} \cdot b_{22} \cdots b_{nn}$.

Suppose that $\pi$ is a prime (mod $p$) and that $\mathfrak{P}$ is the prime ideal divisor associated with $\pi$. Let $N(\mathfrak{P}) = p^f$, i.e., $N(\pi) = p^f q$ where $(p, q) = 1$. We prove the following lemmas:

LEMMA 1. *If* $\omega_i \equiv 0 \pmod{\mathfrak{P}}$ *then* $b_{ii} \not\equiv 0 \pmod{p}$.

For if $\omega_i \equiv 0 \pmod{\mathfrak{P}}$ there exists a $c$, prime to $p$, such that $c\omega_i$ is a multiple of $\pi$. Hence $c \equiv 0 \pmod{b_{ii}}$ and therefore $b_{ii} \not\equiv 0 \pmod{p}$.

LEMMA 2. $b_{ii} \not\equiv 0 \pmod{p^2}$, $i = 1, 2, \cdots, n$.

For since $\pi | p$ there exists a $c$, prime to $p$, such that $cp$, and hence $cp\omega_i$, is a multiple of $\pi$. Hence $cp \equiv 0 \pmod{b_{ii}}$ and therefore $b_{ii} \not\equiv 0 \pmod{p^2}$.

From Lemmas 1 and 2 we obtain the following theorem:

THEOREM 6. *If* $N(\pi) = p^f q$, $(p, q) = 1$, *precisely* $f$ *of the integers* $b_{11}, b_{22}, \cdots, b_{nn}$ *are divisible by* $p$.

Let $\omega_{i_1}, \omega_{i_2}, \cdots, \omega_{i_n}$ be the basis integers for which $b_{ii} \equiv 0 \pmod{p}$ retaining the order in (4). We have the following theorem:

THEOREM 7. *If* $\alpha$ *is a linear combination of* $\omega_{i_1}, \omega_{i_2}, \cdots, \omega_{i_f}$ *with rational coefficients, then*

$$\alpha = c_1\omega_{i_1} + c_2\omega_{i_2} + \cdots + c_f\omega_{i_f} \equiv 0 \pmod{\mathfrak{P}} \tag{5}$$

*if and only if*

$$c_1\omega_{i_1} + c_2\omega_{i_2} + \cdots + c_f\omega_{i_f} \equiv 0 \pmod{p}. \tag{6}$$

For if $\alpha \equiv 0 \pmod{\mathfrak{P}}$ there exists a $c$, prime to $p$, such that $c\alpha$ is a multiple of $\pi$. Hence $c \cdot c_f \equiv 0 \pmod{b_{i_f i_f}}$ and therefore $c_f \equiv 0 \pmod{p}$. Since $\pi | p$ it follows that

$$c_1\omega_{i_1} + c_2\omega_{i_2} + \cdots + c_{f-1}\omega_{i_{f-1}} \equiv 0 \pmod{\mathfrak{P}}.$$

Hence $c_{f-1} \equiv 0 \pmod{p}$. Continuing in this way we obtain (6).

THEOREM 8. *Every integer $\omega$ in $K$ satisfies a congruence of the form*

$$\omega \equiv c_1\omega_{i_1} + c_2\omega_{i_2} + \cdots + c_f\omega_{i_f} \pmod{\mathfrak{P}}. \tag{7}$$

If $b_{ii} \not\equiv 0 \pmod{p}$ we may choose a $d_i$ such that $d_i b_{ii} \equiv 1 \pmod{p}$. Multiplying $\Omega_i$ in (4) by $d_i$ we obtain, since $\Omega_i \equiv 0 \pmod{\mathfrak{P}}$,

$$\omega_i \equiv r_1\omega_1 + r_2\omega_2 + \cdots + r_{i-1}\omega_{i-1} \pmod{\mathfrak{P}}.$$

Hence, by successive elimination of the $\omega_i$'s for which $b_{ii} \not\equiv 0 \pmod{p}$ from the expression for $\omega$ in terms of the basis integers, we obtain (7).

From Theorems 7 and 8 we obtain immediately

THEOREM 9. *The norm of a prime ideal divisor $\mathfrak{P}$ is equal to the number of incongruent residue classes* (mod $\mathfrak{P}$) *of integers in $K$.*

To extend Theorem 9 to any ideal divisor we prove the following lemma:

LEMMA 3. *The number of incongruent classes* (mod $\mathfrak{M}\cdot\mathfrak{N}$) *is equal to the product of the number of classes* (mod $\mathfrak{M}$) *and the number of classes* (mod $\mathfrak{N}$).

For we may determine an integer $\mu$ of $K$ divisible by $\mathfrak{N}$ such that $\mu/\mathfrak{N}$ is prime to $\mathfrak{M}$.* Then, if $\eta_i$ and $\zeta_j$ run over a complete residue system (mod $\mathfrak{M}$) and (mod $\mathfrak{N}$) respectively, the integers $\eta_i\mu + \zeta_j$ form a complete residue system (mod $\mathfrak{M}\cdot\mathfrak{N}$) as may be shown in the usual way. Using Definition 7 we obtain the following theorem:

THEOREM 10. *The norm of an ideal divisor $\mathfrak{M}$ is equal to the number of incongruent classes* (mod $\mathfrak{M}$) *of integers in $K$.*

The theorems on congruences for ideal moduli in Section 8 of Hilbert's report† follow directly.

## IV. THE THEOREM OF DEDEKIND

Let $\theta$ be a primitive integer of $K$ and a root of the equation

$$f(x) = x^n + a_1x^{n-1} + \cdots + a_n = 0. \tag{8}$$

If $\omega_1, \omega_2, \cdots, \omega_n$ is an integral base of $K$ and

$$\theta^{i-1} = a_{1i}\omega_1 + a_{2i}\omega_2 + \cdots + a_{ni}\omega_n \qquad (i = 1, 2, \cdots, n), \tag{9}$$

* The existence of $\mu$ follows from the independence theorem of Tchebotarev, loc. cit., p. 127.

† D. Hilbert, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 4 (1894–95), pp. 191–192.

we have

$$D_\theta = k_\theta^2 d$$

where $D_\theta$ is the discriminant of $\theta$, $d$ is the discriminant of $K$, and $k_\theta = |a_{ij}|$ is the index of $\theta$. By solving (9) for $\omega_1, \omega_2, \cdots, \omega_n$ we obtain the following theorem:

THEOREM 11. *Every integer $\omega$ in $K$ can be expressed in the form*

$$\omega = \frac{b_0 + b_1\theta + \cdots + b_{n-1}\theta^{n-1}}{k_\theta}.$$

The following theorem follows from Definition 3:

THEOREM 12. *If $p$ does not divide $k_\theta$ then every integer of $K$ is associate* (mod $p$) *to a polynomial in $\theta$.*

THEOREM 13. *If $F(x)$ and $G(x)$ are relatively prime* (mod $p$), *then $F(\theta)$ and $G(\theta)$ are relatively prime* (mod $p$).

For if $F(x)$ and $G(x)$ are relatively prime (mod $p$) there exist polynomials $A(x)$ and $B(x)$ such that

$$A(x)F(x) + B(x)G(x) \equiv 1 \pmod{p}$$

and hence

$$A(\theta)F(\theta) + B(\theta)G(\theta) \equiv 1 \pmod{p}.$$

Since all primes (mod $p$) divide $p$ (mod $p$) it follows that any common divisor of $F(\theta)$ and $G(\theta)$ must be a unit (mod $p$).

DEFINITION 8. *By Theorems 3 and 12 we may choose any prime function $\phi(x)$* (mod $p$) *in such a way that $\phi(\theta) | p$. We shall call such prime functions "normal."*

THEOREM 14. *If $p$ does not divide $k_\theta$ and $\phi(x)$ is a normal prime function* (mod $p$) *then $\phi(\theta)$ is either a prime or a unit* (mod $p$).

Every polynomial $F(x)$ is either prime (mod $p$) to $\phi(x)$ or divisible (mod $p$) by $\phi(x)$. In the first case $F(\theta)$ is prime (mod $p$) to $\phi(\theta)$. In the second case $F(\theta)$ is divisible (mod $p$) by $\phi(\theta)$, for we have

$$F(\theta) = \phi(\theta)G(\theta) + pH(\theta),$$

and since $\phi(\theta) | p$, it follows that $\phi(\theta) | F(\theta)$. Hence every integer in $K$ is either divisible by $\phi(\theta)$ or prime to it and, by Definition 2, it follows that $\phi(\theta)$ is either a prime or a unit (mod $p$).

THEOREM 15. *If $p$ does not divide $k_\theta$ and $\phi(x)$ is a normal prime function* (mod $p$) *of $f$th degree, then* $N(\phi(\theta)) \not\equiv 0 \pmod{p^{f+1}}$.

Suppose that $N(\phi(\theta)) \equiv 0 \pmod{p}$. Then $\phi(\theta)$ is not a unit (mod $p$) and hence is a prime (mod $p$). Let $\mathfrak{P}$ be the prime ideal divisor associated with the prime divisor $\phi(\theta)$. We have

$$\phi(\theta) = \theta^f + c_1\theta^{f-1} + \cdots + c_f \equiv 0 \pmod{\mathfrak{P}}.$$

Hence, from Theorem 12, any integer $\omega$ in $K$ satisfies a congruence of the form

$$\omega \equiv b_1\theta^{f-1} + b_2\theta^{f-2} + \cdots + b_f \pmod{\mathfrak{P}}.$$

Since $p \equiv 0 \pmod{\mathfrak{P}}$, it follows that the number of incongruent classes of integers in $K$ is less than or equal to $p^f$, i.e., $N(\mathfrak{P}) \leqq p^f$. Hence, by Definition 6, $N(\phi(\theta)) \not\equiv 0 \pmod{p^{f+1}}$.

We are now prepared to prove the anologue of Dedekind's theorem:

THEOREM 16. *If $p$ does not divide $k_\theta$ and*

$$f(x) \equiv \phi_1(x)^{e_1}\phi_2(x)^{e_2} \cdots \phi_r(x)^{e_r} \pmod{p}$$

*where $\phi_i(x)$ is a normal prime function* (mod $p$) *of degree $f_i$, $i = 1, 2, \cdots, r$, then*

$$cp = \epsilon \cdot \phi_1(\theta)^{e_1} \cdot \phi_2(\theta)^{e_2} \cdots \phi_r(\theta)^{e_r}, \tag{10}$$

*where $(c, p) = 1$ and $\epsilon$ is a unit* (mod $p$), *is the decomposition of $p$ in distinct prime divisors* (mod $p$).

For let

$$f(x) = \phi_1(x)^{e_1}\phi_2(x)^{e_2} \cdots \phi_r(x)^{e_r} + pM(x).$$

Then

$$-pM(\theta) = \phi_1(\theta)^{e_1}\phi_2(\theta)^{e_2} \cdots \phi_r(\theta)^{e_r}.$$

Taking the norm of both sides we have

$$-p^n N(M(\theta)) = N(\phi_1(\theta))^{e_1} N(\phi_2(\theta))^{e_2} \cdots N(\phi_r(\theta))^{e_r}.$$

Since $\sum_{i=1}^{r} e_i f_i = n$, it follows from Theorem 15 that $N(M(\theta)) \not\equiv 0 \pmod{p}$ and furthermore $N(\phi_i(\theta)) \equiv 0 \pmod{p^{f_i}}$. Hence $\phi_i(\theta)$, $i = 1, 2, \cdots, r$, is a prime (mod $p$) and furthermore, by Theorem 13, $\phi_i(\theta)$ and $\phi_j(\theta)$ are distinct (mod $p$) for $i \neq j$. Also, $M(\theta)$ is a unit (mod $p$) and there exists a $c$, prime to $p$, such that $c/M(\theta) = \epsilon$ is a unit (mod $p$) and (10) follows.

Expressed in terms of the ideal divisors of Zolotarev, Theorem 16 becomes

THEOREM 17. *Under the hypothesis of Theorem* 16,

$$p = \mathfrak{P}_1^{e_1}\mathfrak{P}_2^{e_2}\cdots\mathfrak{P}_r^{e_r},\quad N(\mathfrak{P}_i) = p^{f_i},$$

*where* $\mathfrak{P}_i$ *is a prime ideal divisor associated with the normal prime divisor* $\phi_i(\theta) \pmod p$, $i=1, 2, \cdots, r$.

To express this theorem in the form of Theorem 1 we must remove the condition of normalcy of the divisors $\phi_i(x)$. Suppose that

$$f(x) \equiv \prod_{i=1}^{r}\Phi_i(x)^{e_i} \pmod p$$

is a decomposition of $f(x)$ satisfying the conditions of Theorem 1. By Theorems 3 and 12 there exists a prime function decomposition

$$f(x) \equiv \prod_{i=1}^{r}\phi_i(x)^{e_i} \pmod p$$

in which the $\phi_i(x)$ are all normal and

$$\phi_i(x) \equiv \Phi_i(x) \pmod{p;\ i = 1, 2, \cdots, r}. \tag{11}$$

This may also be shown directly as follows. Suppose

$$f(x) = \prod_{i=1}^{r}\Phi_i(x)^{e_i} + pM(x).$$

Since we have supposed that $p$ is not an index divisor it follows by a criterion due to Dedekind* that if $e_i>1$ then $M(x)\not\equiv 0 \pmod{p, \Phi_i(x)}$. Now let

$$P_i(x) = \left[\prod_{j=1}^{r}\Phi_j(x)^{e_i}\right] \Big/ \Phi_i(x) \qquad (i = 1, 2, \cdots, r).$$

We distinguish two types of $\Phi(x)$, (1) those for which $M(x)\equiv 0 \pmod{p, \Phi_i(x)}$ and (2) those for which $M(x)\not\equiv 0 \pmod{p, \Phi_i(x)}$. If $\Phi_i(x)$ is of type (1) we write $\phi_i(x)=\Phi_i(x)+pP_i(x)$, if $\Phi_i(x)$ is of type (2) we write $\phi_i(x)=\Phi_i(x)$. Then

$$f(x) = \prod_{i=1}^{r}\phi_i(x)^{e_i} + pM'(x), \tag{12}$$

where $M'(x)\equiv -[P_i(x)]^2 \pmod{p, \Phi_i(x)}$ in case (1) and $M'(x)\equiv M(x) \pmod{p, \Phi_i(x)}$ in case (2). Hence $M'(x)\not\equiv 0 \pmod{p, \Phi_i(x)}$, $i=1, 2, \cdots r$,

* Cf. P. Bachmann, *Allgemeine Arithmetik der Zahlenkörper*, Leipzig, 1926, p. 277.

i.e., $M'(x) \not\equiv 0 \pmod{p, \phi_i(x)}$. Setting $x=\theta$ in (12) we see that $\phi_i(\theta)^{e_i} | p$, $i=1, 2, \cdots, r$, i.e., all the $\phi_i(x)$ are normal.

If $e_i>1$, it follows from (11) that $\Phi_i(\theta)$ is divisible (mod $p$) by precisely the first power of $\phi_i(\theta)$ and by no other divisors of $p$. Hence we may write $\phi_i(\theta)$ as the greatest common divisor (mod $p$) of $p$ and $\Phi_i(\theta)$; symbolically, $\mathfrak{P}_i=(p, \Phi_i(\theta))$. If $e_i=1$, then $p$ will contain precisely the first power of $\phi_i(\theta)$ and again $\mathfrak{P}_i=(p, \Phi_i(\theta))$. Hence we have Theorem 1.

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIF.

# POLYNOMIALS OF BEST APPROXIMATION ON AN INFINITE INTERVAL*

BY

JAMES M. EARL

**Introduction.** This paper is concerned with an extension of certain studies of the degree of convergence of methods of approximation to a given function. The approximating functions are polynomials determined so as to minimize the integral over an infinite interval of the product of a non-negative weight function and the $m$th power of the magnitude of the error. The essential point of this discussion is that the interval of integration is infinite. The case of a finite interval has been considered by Jackson† and others for polynomials as well as for other types of approximating functions.

The problem may be formulated more precisely as follows. Let $f(x)$ be a given function of $x$ defined on an infinite interval $S$. Under suitable restrictions on $f(x)$ and the non-negative weight function $r(x)$, there exists a determination of the coefficients in the polynomial

$$p_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for which the integral

$$\int_S r(x) \mid f(x) - p_n(x) \mid^m dx \tag{1}$$

is a minimum. When the coefficients in $p_n(x)$ are so determined, the polynomial $p_n(x)$ is called an approximating polynomial of degree $n$ and is denoted by $P_{mn}(x)$ while the value of the minimized integral is denoted by $I_{mn}$. The problem is to investigate the degree of the uniform convergence of $P_{mn}(x)$ to $f(x)$ when $m$ is a fixed positive number and $n$ is allowed to become infinite.

The Hermite series and the Laguerre series are two important special cases of the least-square problem, $m=2$, on the doubly infinite interval $(-\infty, \infty)$ and the simply infinite interval $(0, \infty)$ respectively. The former corresponds to the weight function $r(x)=\exp(-x^2)$ and the latter to $r(x)=\exp$

* Presented to the Society, April 7, 1928; received by the editors in May, 1930.

† See D. Jackson, *On the convergence of certain trigonometric and polynomial approximations*, these Transactions, vol. 22 (1921), pp. 158–166; *Note on the convergence of weighted trigonometric series*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 259–263; and bibliographical references there noted.

$(-x)$. In these special cases, the problem of uniform convergence has been extensively studied by various methods.* The method used here consists of obtaining an expression which gives an upper bound for the error in the approximation to $f(x)$ in terms of the integral $I_{mn}$ and then of obtaining upper bounds for $I_{mn}$ to be substituted in this expression. By this method, results on uniform convergence are obtained for very general exponential weight functions, for the doubly infinite as well as for the simply infinite interval, and for any given positive $m$. While these results possess considerable generality, they are intended to suggest rather than to exhaust the wide range of applicability of the method.

The arrangement of the paper is in summary as follows. After a statement of well known theorems for reference in §1, the questions of the existence and the uniqueness of the approximating polynomials are considered in §2. The upper bounds for the magnitude of the error $f(x)-P_{mn}(x)$ obtained in §3 involve $I_{mn}$; the question of degree of convergence is thus reduced to the problem of determining upper bounds for the integral $I_{mn}$. This is done in §4 for the case of an essentially even weight function on the doubly infinite interval; the upper bound for $I_{mn}$ obtained depends not only on the nature of $f(x)$ but on the behavior of $r(x)$ for large values of $|x|$. Application to the special case of an exponential weight function $r(x)$ is made in §5. The case of a simply infinite interval taken for definiteness as $(0, \infty)$ is made to depend on the above in §6 and it is found that the necessary restriction on $r(x)$ for large values of $|x|$ is less for the simply than for the doubly infinite interval. Section 7 contains a method for the treatment of essentially skew weight functions on the doubly infinite interval. It is shown in the application to an exponential weight function that the necessary restriction on $r(x)$ for large positive values of $x$ can be weakened at the expense of increasing the restriction on $r(x)$ for large negative values of $x$. Unification of the results of §§3 to 6 is incidentally effected in §7.

1. **Theorems for reference.** The theorems due to S. Bernstein,†A. Markoff,‡

---

* For bibliography, see E. Hille, *A class of reciprocal functions*, Annals of Mathematics, (2) vol. 27 (1926), pp. 427–464. See also W. E. Milne, *On the degree of convergence of the Gram-Charlier series*, these Transactions, vol. 31 (1929), pp. 422–433; the concluding section of Chapter III in the colloquium lectures of D. Jackson, *The Theory of Approximation*, New York, 1930; and an abstract by J. M. Earl, *The degree of convergence of Hermite and Laguerre series*, Bulletin of the American Mathematical Society, vol. 35 (1930), p. 455.

† S. Bernstein, *Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degré donné*, Mémoire couronné, Brussels, 1912.

‡ For proof see M. Riesz, *Eine trigonometrische Interpolationsformel* · · · , Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 23 (1914), pp. 354–368.

P. L. Tchebycheff,* and D. Jackson,† respectively, that follow are lettered for convenience of reference.

In Theorems B, M and T below, let $q_n(x)$ with derivative $q_n'(x)$ be any polynomial of degree $n$ which satisfies the stated hypothesis.

THEOREM B. *If* $|q_n(x)| \leqq L$ *on* $(a, b)$, *then* $|q_n'(x)| \leqq nL/[(b-x)\ (x-a)]^{1/2}$ *on* $(a, b)$.

THEOREM M. *If* $|q_n(x)| \leqq L$ *on* $(a, b)$, *then* $|q_n'(x)| \leqq 2n^2L/(b-a)$ *on* $(a, b)$.

THEOREM T. (a) *If* $|q_n(x)| \leqq L$ *on* $(-a, a)$, *then* $|q_n(x)| \leqq L|2x/a|^n$ *when* $x$ *is not on* $(-a, a)$.

(b) *If* $|q_n(x)| \leqq L$ *on* $(0, a)$ [*or on* $(-a, 0)$], *then* $|q_n(x)| \leqq L|4x/a|^n$ *on* $(a, \infty)$ [*or on* $(-\infty, -a)$].

THEOREM J. *If on the interval* $(a, b)$, $f(x)$ *has a pth derivative* $f^{(p)}(x)$ *which satisfies a Lipschitz condition of order* $u$ *with coefficient* $L$, *that is*

$$|f(x_2) - f(x_1)| \leqq L|x_2 - x_1|^u \quad (0 < u \leqq 1)$$

*for every pair of values* $x_1$, $x_2$ *on* $(a, b)$, *then for each integer* $n > p$, *there exists a polynomial* $p_n(x)$ *of degree* $n$ *such that*

$$|f(x) - p_n(x)| \leqq KL[(b-a)/n]^{p+u}$$

*uniformly on* $(a, b)$ *where* $K$ *is an absolute constant.*

2. **The existence and the uniqueness of the approximating polynomials.** The first lemma below may be used as the basis for proofs of existence of approximating functions in a variety of least $m$th power problems. A similar‡ lemma has been proved for bounded§ $h$'s. The functions $h_1(x), \cdots, h_n(x)$ are called properly independent on a point set $S$ if every linear combination of these functions in which the coefficients are not all zero is different from zero on a sub-set of $S$ of positive measure. The set $S$ may be thought of as bounded or not bounded in this section.

---

* For proof see S. Bernstein, *Leçons sur les Propriétés Extrêmales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle*, Paris, 1926, pp. 7–8.

† See, for example, D. Jackson, *On the approximate representation of an indefinite integral* $\cdots$, these Transactions, vol. 14 (1913), Theorem VIII, p. 354.

‡ D. Jackson, *A generalized problem in weighted approximation*, these Transactions, vol. 26 (1924), pp. 133–154; pp. 133–137.

§ Shohat has removed the restriction of boundedness in a special case for polynomial approximation; see J. Shohat, *On the polynomial of best approximation to a given continuous function*, Bulletin of the American Mathematical Society, vol. 31 (1925), pp. 509–514; p. 510.

LEMMA 1. *Suppose that the functions* $h_1(x), \cdots, h_n(x)$ *are properly independent on* $S$ *and that the integral*

$$J = \int_S | c_1h_1(x) + \cdots + c_nh_n(x) |^m dx$$

*exists for each set of coefficients* $(c_1, \cdots, c_n)$. *Then there exists a positive constant* $A$ *which is independent of the* $c$'*s such that*

$$| c_k | \leqq (J/A)^{1/m} \qquad (k = 1, \cdots, n). \tag{2}$$

The generality of the hypothesis that the integral $J$ exists for each set of coefficients is apparent on noticing that this hypothesis is satisfied for example when each $h$ is measurable and has a summable $m$th power on $S$. The proof of Lemma 1 follows.

Let $C$ denote the set of points $(c_1, \cdots, c_n)$ in $n$ dimensions such that $c_k=1$ for some $k$ and $|c_k| \leqq 1$ for all $k=1, \cdots, n$. Corresponding to each point of $C$, the integral $J$ is positive. The set $C$ is closed. Therefore since $J$ is a continuous function of $c_1, \cdots, c_n$, bounded for all points of $C$, there exists a least $J=A$ which is positive.

Now let the $c$'s in $J$ be arbitrary. The lemma is true if the $c$'s are all zero. Suppose then that some $c$, say $c_k$, is not zero and has at least as great magnitude as any other coefficient. Then the coefficients in the quotient $[c_1h_1(x) + \cdots + c_nh_n(x)]/c_k$ are a point of $C$ so that $J/|c_k|^m \geqq A$ and from this follows the inequality (2).

THEOREM 1. *Suppose that for each polynomial* $p_n(x)$ *of degree* $n$, *the integral* (1) *exists and that* $\int_S r(x)dx = R > 0$.

(a) *If* $m$ *is positive, at least one approximating polynomial* $P_{mn}(x)$ *exists.*

(b) *If* $m$ *exceeds unity, the approximating polynomial* $P_{mn}(x)$ *is unique.*

(c) *If* $m=1$ *and* $f(x)$ *is continuous, the approximating polynomial* $P_{mn}(x)$ *is unique.*

Statement (a) of this theorem follows from Lemma 1 by the usual reasoning. The usual proof of (b) applies here while (c) may be proved by extending a proof given by Shohat* for the case of a finite interval $S$. These proofs are omitted.

3. **The degree of convergence of the approximating polynomials.** Here $(a, b)$ denotes a finite interval which is a sub-set of the set $S$ which may be thought of as bounded or not bounded. If $r(x)$ satisfies Hypothesis Y (given below) on this interval, then upper bounds in terms of $n$ and $I_{mn}$ for $|f(x) - P_{mn}(x)|$ on $(a, b)$ and $(a+\epsilon, b-\epsilon)$ are given below by expressions (9) and (10)

* J. Shohat, loc. cit., pp. 511–513.

respectively where $\epsilon$ is any preassigned positive constant less than $\frac{1}{2}(b-a)$. In the proofs, $U$ denotes the least upper bound of $|f(x)|$ on $(a, b)$ and it is assumed that $|P_{mn}(x)|$ attains its maximum value $\pi_j$ on $(a+j\epsilon/k, b-j\epsilon/k)$ at $x=x_j$ $(j=0, 1, \cdots, k)$, $k$ being any given positive integer. Various constants which are independent of $n$ are denoted by the single letter $c$. The minimized integral $I_{mn}$ is of course taken over the whole of $S$.

LEMMA 1. *If $r(x)$ has a positive lower bound $v$ on $(a, b)$ and if $\pi_j > 4U$, then*

$$\pi_j \leqq c(nI_{mn}\pi_{j-1})^{1/(m+1)} \qquad (j = 1, \cdots, k). \tag{3}$$

By Theorem B, $|P'_{mn}(x)| \leqq (kn/\epsilon)\pi_{j-1}$ on $(a+j\epsilon/k, b-j\epsilon/k)$. It follows by the law of the mean that

$$|P_{mn}(x) - P_{mn}(x_j)| \leqq (kn/\epsilon)\pi_{j-1}|x - x_j| \leqq \tfrac{1}{2}\pi_j$$

whenever $|x-x_j| \leqq \pi_j(2kn\pi_{j-1}/\epsilon)^{-1}$ so that $|P_{mn}(x)| \geqq \frac{1}{2}\pi_j$ on an interval $w$ of length $c\pi_j/(n\pi_{j-1})$ which lies wholly on $(a+j\epsilon/k, b-j\epsilon/k)$ and has $x_j$ for one end point. It follows that

$$|f(x) - P_{mn}(x)| \geqq |P_{mn}(x)| - |f(x)| \geqq \tfrac{1}{2}\pi_j - \tfrac{1}{4}\pi_j = \tfrac{1}{4}\pi_j$$

on $w$ and, since $r(x) \geqq v$ on $w$, that

$$I_{mn} \geqq c\pi_j^m \int_w r(x)dx \geqq c\pi_j^{m+1}/(n\pi_{j-1}). \tag{4}$$

The inequality (3) of the lemma results on solution of this inequality for $\pi_j$.

**Hypothesis Y.** The function $r(x)$ is said to satisfy Hypothesis Y on $(a, b)$ if there exist positive constants $D$, $d$ and non-negative constants $g$, $s$ such that $r(x)$ has a positive lower bound $v=Dd^{g+s}$ on $(a, b)$ except perhaps on $(l+2)$ non-overlapping sub-intervals $(a, a+d)$, $(y_i-d, y_i+d)$ $(i=1, \cdots, l)$ and $(b-d, b)$ on which $r(x)$ behaves as follows:

$$r(x) \geqq D(b-x)^g \quad \text{whenever } b-x \leqq d,$$
$$r(x) \geqq D(x-a)^g \quad \text{whenever } x-a \leqq d,$$
$$r(x) \geqq D|x-y_i|^s \text{ whenever } |x-y_i| \leqq d.$$

Under Hypothesis Y, the weight function may become zero in a restricted manner at a finite number of interior points as well as at the end points of $(a, b)$. If, for example, $S$ is the interval $(0, \infty)$ and $(a, b)$ is $(0, A)$, then using the upper bound (9) of Theorem 2 for the error, it is seen that there is no weakening in the degree of convergence provable caused by allowing the weight function to become zero at a finite number of interior points of $(0, A)$ provided the order of vanishing in no case exceeds $s=1$. The essential novelty of the convergence proofs given here lies in the difference in the effect of

allowing $r(x)$ to vanish to a given order at interior points of $(a, b)$ as compared with the effect of allowing $r(x)$ to vanish to the same order at one or both end points of $(a, b)$.

In Lemmas 2 and 4 below, $(\alpha, \beta)$ denotes any given interval which is interior to $(a, b)$ and which contains the intervals $(y_i-d, y_i+d)$ $(i=1, \cdots, l)$.

LEMMA 2. *Suppose that $r(x)$ satisfies Hypothesis* Y *on $(a, b)$ and that $\pi_0>4U$. Then*

$$\pi_0 \leqq c(n^{s+1}I_{mn})^{1/m} \text{ if } x_0 \text{ is on } (\alpha, \beta), \tag{5}$$

$$\pi_0 \leqq c(n^{2(s+1)}I_{mn})^{1/m} \text{ if } x_0 \text{ is on } (a, \alpha) \text{ or on } (\beta, b). \tag{6}$$

If $x_0$ is on $(\alpha, \beta)$, then by Theorem B, $|P'_{mn}(x)| \leqq cn\pi_0$ on $(\alpha, \beta)$. Steps similar to those used in the proof of Lemma 1 except that $w$ is here replaced by $w'$ of length $1/(2cn)$ with $x_0$ as one end point and that $r(x)$ which is replaced by $v$ in relation (4) is here replaced by $D|x_0-x|^s$, result in inequality (5).

If $x_0$ is on $(a, \alpha)$ [or on $(\beta, b)$], Theorem M gives $|P'_{mn}(x)| \leqq cn^2\pi_0$. The proof of (6) is similar to the proof of (5) but $w'$ is here replaced by $w''$ of length $1/(2cn^2)$ and $s$ is replaced by $g$.

The following lemma results on setting $s=g=0$ in inequalities (5) and (6).

LEMMA 3. *If $r(x)$ has a positive lower bound on $(a, b)$ and if $\pi_0>4U$, then $\pi_0 \leqq c(n^2I_{mn})^{1/m}$.*

LEMMA 4. *Suppose that $r(x)$ satisfies Hypothesis* Y *on $(a, b)$ and that $\pi_k>4U$. Then*

$$\pi_k \leqq c(n^{s+1}I_{mn})^{1/m} \tag{7}$$

*if $x_k$ is on $(\alpha, \beta)$,*

$$\pi_k \leqq c(n^HI_{mn})^{1/m}, \tag{8}$$

*where $H=1+(m+1)^{-k}$, if $x_k$ is on $(a+\epsilon, \alpha)$ or on $(\beta, b-\epsilon)$.*

If $x_k$ is on $(\alpha, \beta)$, inequality (7) is obtained in the same manner as (5) of Lemma 2. If $x_k$ is on $(a+\epsilon, \alpha)$, then when $\epsilon$ is sufficiently small, $r(x)$ has a positive lower bound on $(a+\frac{1}{2}\epsilon, \alpha+\frac{1}{2}\epsilon)$. Let $\pi_0'$ be the maximum of $|P_{mn}(x)|$ on this interval. By Lemma 3, $\pi_0' \leqq c(n^2I_{mn})^{1/m}$ so that the inequality (8) results by successive applications of Lemma 1 with $(a, b)$ replaced by $(a+\frac{1}{2}\epsilon, \alpha+\frac{1}{2}\epsilon)$ and $\epsilon$ by $\frac{1}{2}\epsilon$. This establishes Lemma 4.

If the magnitude of the error in the approximation to $f(x)$ on $(a, b)$ by a polynomial $q_n(x)$ does not exceed $\epsilon_n$, then $U$ for the function $f(x)-q_n(x)$ does

not exceed $\epsilon_n$ and, since, by Lemma 2, the approximating polynomial $P_{mn}(x) - q_n(x)$ for this function does not exceed $4\epsilon_n$ plus the sum of the expressions on the right in inequalities (5) and (6), the magnitude of $f(x)-P_{mn}(x)$ does not exceed the expression (9) on $(a, b)$. Similar reasoning using Lemma 4 instead of Lemma 2 results in the upper bound (10) of Theorem 2 for the error on $(a+\epsilon, b-\epsilon)$.

THEOREM 2. *Suppose that $f(x)$ can be uniformly approximated by a polynomial of degree $n$ with an error which nowhere exceeds $\epsilon_n$ on the interval $(a, b)$. Let $\epsilon$ be any positive number less than $\frac{1}{2}(b-a)$. If $r(x)$ satisfies Hypothesis* Y *on $(a, b)$, then $|f(x)-P_{mn}(x)|$ does not exceed*

$$5\epsilon_n + c[(n^{2(\sigma+1)} + n^{\sigma+1})I_{mn}]^{1/m} \text{ on } (a, b), \tag{9}$$

$$5\epsilon_n + c[(n^H + n^{\sigma+1})I_{mn}]^{1/m} \text{ on } (a+\epsilon, b-\epsilon), \tag{10}$$

*where $H=1+(m+1)^{-k}$.*

**4. The order of magnitude of the minimized integral.** An upper bound for the integral $I_{mn}$ is obtained here for the case of an even weight function $r(x)$. Extensions to the case of more general weight functions are then suggested. The set $S$ is taken to be the doubly infinite interval on which $r(x)$, as well as the product of $r(x)$ and the $m$th power of each of the functions $x^\alpha M(x)$ $(\alpha>0)$ and $|f(x)-p_n(x)|$, is assumed to be summable where $p_n(x)$ is any polynomial of degree $n$ and $M(x)$ is even, non-negative, non-decreasing for non-negative values of $x$ and such that on every interval $(-a, a)$, the $p$th derivative $f^{(p)}(x)$ of $f(x)$, assumed to exist, satisfies a Lipschitz condition of order $u$ with coefficient $M(a)$. In addition to the above, let it be assumed throughout that $f(x)$ and its first $p$ derivatives are zero when $x=0$; this condition may always be realized by adding to $f(x)$ a polynomial of degree $p$ and since the approximations to $f(x)$ are by polynomials of degree $n>p$, there is no loss of generality in the assumption.

It will be convenient to denote by $G_{Mr}(a, A)$, $H_{Mr}(a, A)$, $K_n$, and $V_p(x, \alpha)$ the following expressions where $A \geqq a>0$ and where $q_n(x)$ denotes a polynomial of degree $n$:

$$G_{Mr}(a, A) = [M(a)A^{p+u}]^m\left[n^{-m(p+u)} + \int_A^\infty r(x)(2x/A)^{mn}dx\right],$$

$$H_{Mr}(a, A) = G_{Mr}(a, A) + \int_a^\infty r(x)[M(x)x^{p+u}]^m dx,$$

$$K_n = \int_{-\infty}^\infty r(x)\,|f(x) - q_n(x)|^m dx,$$

$$V_p(x, \alpha) = f(\alpha) + f'(\alpha)(x-\alpha) + \cdots + f^{(p)}(\alpha)(x-\alpha)^p/p!.$$

The various constants denoted hereafter by $c$ are independent not only of $n$ but of $A$, $a$ and $M(x)$.

In Lemmas 1 to 3, let it be assumed that $r(x)$ is an even function of $x$. The following lemma results by successive integrations from 0 to $x$.

LEMMA 1. *For every real value of* $x$, $|f^{(p-j)}(x)| \leqq M(x)|x|^{j+u}$ $(j=0, 1, \cdots, p)$.

LEMMA 2. *If* $M(x)=M$ *is constant, then for every integer* $n>p$, *there exists a polynomial* $q_n(x)$ *of degree* $n$ *such that* $K_n$ *does not exceed* $cG_{M_r}(a, A)$.

By Theorem J, there exists a polynomial $q_n(x)$ such that

$$|f(x) - q_n(x)| \leqq cM(A/n)^{p+u} \tag{11}$$

on $(-A, A)$. Moreover, $|f(x)|$ does not exceed $MA^{p+u}$ on $(-A, A)$, by Lemma 1, so that for points on $(-A, A)$ by (11) and therefore for points not on $(-A, A)$ by Theorem T, $|q_n(x)|$ does not exceed $cMA^{p+u}$ and $cMA^{p+u}|2x/A|^n$ respectively. By Lemma 1, $|f(x)|$ nowhere exceeds $MA^{p+u}|x/A|^{p+u}$. Therefore, since $(p+u)$ does not exceed $n$,

$$|f(x) - q_n(x)| \leqq cMA^{p+u}|2x/A|^n \tag{12}$$

when $x$ is not on $(-A, A)$ because then $|x/A|$ exceeds unity. Since $r(x)$ is even, the lemma results on replacing the error $|f(x)-q_n(x)|$ in $K_n$ by its upper bounds (11) and (12) on and not on $(-A, A)$ respectively.

In the following lemma, the restriction that $M(x)$ be constant is removed.

LEMMA 3. *For every integer* $n>p$, *there exists a polynomial* $q_n(x)$ *of degree* $n$ *such that* $K_n$ *does not exceed* $cH_{M_r}(a, A)$.

The auxiliary function $g(x)$ which is defined as $f(x)$, $V_p(x, a)$ and $V_p(x, -a)$ when $x$ is on $(-a, a)$, $(a, \infty)$ and $(-\infty, -a)$ respectively satisfies the conditions of Lemma 2 with $M=M(a)$, so that there exists a polynomial $q_n(x)$ such that the first integral on the right in the inequality

$$K_n \leqq 2^m \int_{-\infty}^{\infty} r(x)|g(x) - q_n(x)|^m dx + 2^m \int_{-\infty}^{\infty} r(x)|f(x) - g(x)|^m dx \tag{13}$$

does not exceed $cG_{M_r}(a, A)$. Moreover, by Lemma 1, neither $|f(x)|$ nor $|g(x)|$ exceeds $M(x)|x|^{p+u}$ and therefore since the integrand in the second integral on the right in (13) is zero on $(-a, a)$, this integral does not exceed a constant multiple of the last term in the expression for $H_{M_r}(a, A)$. This establishes Lemma 3.

The following theorem is an immediate consequence of Lemma 3 because the integral $I_{mn}$ never exceeds $K_n$.

THEOREM 3. *If $S$ is the doubly infinite interval and if $r(x)$ is an even function of $x$, then for every integer $n > p$,*

$$I_{mn} \leqq cH_{Mr}(a, A). \tag{14}$$

The significance of the various parameters entering into the somewhat complicated expression $H_{Mr}(a, A)$ will be made clearer as the applications of the above theorem are worked out in detail.

Now, to what related classes of weight functions may the theory of this section be extended? This is a question which naturally arises because it is sometimes desirable to use results for a simple class of weight functions in obtaining results for other classes of weight functions. No complete answer to the question is given here. It may be pointed out, however, that the upper bound for $I_{mn}$ in Theorem 3 is not increased as to order of magnitude when $r(x)$ is replaced by a constant multiple of $r(x)$, or more generally by the product of $r(x)$ and a non-negative bounded function of $x$; the convergence of the error $f(x) - P_{mn}(x)$ toward zero, on the other hand, may be retarded when the new weight function vanishes to a higher order than the original weight function $r(x)$. In the proof of Corollary 3 of the next section, it is noted that the magnitude of $I_{mn}$ is essentially unchanged even when the exponential weight function $r(x)$ is replaced by $|\pi(x)|r(x)$ where $\pi(x)$ is any given polynomial. While the problem of the extension of simple classes of weight functions is important, the case of an arbitrary weight function may be treated independently as in §7 of this paper.

5. **The exponential weight function.** Application of §4 to particular functions $r(x)$ and $M(x)$ involves two important steps illustrated respectively by Lemmas 1 and 2 below. In the first step, $A$ in the expression for $H_{Mr}(a, A)$ is chosen as a function of $n$, the choice depending only on $r(x)$. This, by Theorem 3, gives an upper bound for $I_{mn}$ in terms of $n$ and $a$ which is chosen in the second step as a function of $n$. The upper bound for $I_{mn}$ thus obtained may then be substituted for $I_{mn}$ in Theorem 2, giving, as in Theorem 4 below, an upper bound for the error in the approximation to $f(x)$ by $P_{mn}(x)$ in terms of $n$.

The discussion of this section refers to the weight function $r(x) = \exp(-h|x|^\sigma)$ $(h > 0, \sigma > 1)$ and to $M(x) = \exp(j|x|^\sigma)$ $(0 \leqq mj < h)$, except where explicit notation to the contrary is made. Notice that the exponent of $|x|$ in both $r(x)$ and $M(x)$ is the same number $\sigma$; Corollaries 1 and 3 state results which apply to $M(x) = \exp(d|x|^i)$ $(d \geqq 0)$ where $i$ is less than $\sigma$.

LEMMA 1. *The number $A$ can be chosen as a constant multiple of $n^{1/\sigma}$ so that*

$$H_{M_r}(a, A) \leqq c\left[M(a)n^{-(\sigma-1)(p+u)/\sigma}\right]^m + \int_a^\infty \exp(-hx^\sigma)\left[M(x)x^{p+u}\right]^m dx. \tag{15}$$

The integral from $A$ to $\infty$ in the expression for $H_{M_r}(a, A)$ does not exceed the integral of $\exp(-\frac{1}{2}hx^\sigma)$, which is bounded, times the product of $(2/A)^{mn}$ and the maximum $(2mn/(he\sigma))^{mn/\sigma}$ of $x^{mn}\exp(-\frac{1}{2}hx^\sigma)$ on $(0, \infty)$. The lemma therefore results on setting $A$ equal to $h^{1/\sigma}$ times the constant $4(2m/(he\sigma))^{1/\sigma}$.

LEMMA 2. *If $\omega > 1$ is assigned, then $I_{mn}$ does not exceed*

$$cn^{-m(\sigma-1)(p+u)(1-mj/h)/(\omega\sigma)} \tag{16}$$

*provided $n$ is sufficiently great.*

Let $v > 1$ be chosen so that

$$c(n^{m^2j(v-1)(\sigma-1)(p+u)/(h\sigma)} + 1)n^{-m(\sigma-1)(p+u)(1-mj/h)/\sigma} \tag{17}$$

is less than one half the expression (16). The lemma follows from Theorem 3 on noticing that the expression on the right in (15) does not exceed (17) and therefore (16), provided $a = [(v/h)(m/\sigma)(\sigma-1)\cdot(p+u)\log n]^{1/\sigma}$ and provided $n$ is so great that the integral from $a$ to infinity in (15) does not exceed $\exp(-[h-mj]a^\sigma/v)$ which does not exceed one half the expression (16).

THEOREM 4. *Let $\rho$ be any assigned positive number less than 1 and let $(\alpha, \beta)$ be any given finite interval. Then for $n$ sufficiently great, the error $|f(x) - P_{mn}(x)|$ does not exceed*

$$cn^{-\rho(p+u)(1-mj/h)(\sigma-1)/\sigma+1/m} \text{ on } (\alpha, \beta). \tag{18}$$

Let the interval $(a, b)$ in Theorem 2 be $(\alpha-\epsilon, \beta+\epsilon)$; then $g = s = 0$. Under the hypothesis on $f(x)$, $5\epsilon_n$ may be taken, by virtue of Theorem J, as $cn^{-(p+u)}$, so that Theorem 4 results on substitution of the expression (16) for $I_{mn}$ in (10) with $(\omega-1)$ and $(1/k)$ taken sufficiently small.

It is perhaps worth while to dwell at some length on the implications of Theorem 4; the specializations and extensions of this theorem included below are suggestive of the wide range of applicability of the methods of this paper.

Notice first that uniform convergence results whenever the exponent of $n$ in (18) is negative; it is necessary for this that $\sigma$ exceed 1. When $m$ is positive and $\sigma$ does exceed 1, then uniform convergence does always result provided $f(x)$ is restricted by taking $p$ sufficiently great. It will be seen in the next section that when the doubly infinite interval is replaced by a simply

infinite interval, then it is necessary only that $\sigma$ exceed $\frac{1}{2}$ instead of 1; this difference in restriction on $\sigma$ is compromised, in §7, in the case of a skew weight function.

Next consider the most simple case, $j=0$, in which $f^{(p)}(x)$ satisfies an ordinary Lipschitz condition on $(-\infty, \infty)$ with a constant coefficient. Under these circumstances, a sufficient condition for uniform convergence is that $(p+u)(\sigma-1)/\sigma>1/m$. This is also a sufficient condition for uniform convergence when $f^{(p)}(x)$ is considerably less restricted for large values of $|x|$ by taking $M(x)=\exp(d|x|^i)$ $(d\geqq 0,\ i<\sigma)$; this results because the expression (18) need not be altered in form when $M(x)$ is replaced by a new function which nowhere exceeds a constant multiple of $M(x)$. In particular, if $j>0$ is assigned, then there exists a positive constant $\gamma$ such that the new function $M(x)=\exp(d|x|^i)$ $(d\geqq 0,\ i<\sigma)$ nowhere exceeds $\gamma\exp(j|x|^\sigma)$. In fact, the following corollary results when $\rho<1$ and $j>0$ in (18) are assigned so that $\rho(1-mj/h)=\phi$.

COROLLARY 1. *Suppose that* $r(x)=\exp(-h|x|^\sigma)$ *and that* $M(x)=\exp(d|x|^i)$ $(d\geqq 0,\ i<\sigma)$. *Let* $\phi$ *be any assigned positive number less than* 1 *and let* $(\alpha, \beta)$ *be any given finite interval. Then for* $n$ *sufficiently great, the error* $|f(x)-P_{mn}(x)|$ *does not exceed* $cn^{-\phi(p+u)(\sigma-1)/\sigma+1/m}$ *on* $(\alpha, \beta)$.

The following corollary restates the results of Theorem 4 and Corollary 1 for the special case of the Hermite series in which $r(x)=\exp(-x^2)$ and $m=2$.

COROLLARY 2. *Let* $\rho$ *be any assigned positive number less than* 1. *Then the error in the approximation to* $f(x)$ *on any given finite interval by the* $n$*th partial sum of the Hermite series does not exceed*

$$cn^{-(1/2)\rho(p+u)(1-2j)+1/2} \text{ when } M(x)=\exp(jx^2) \qquad (0<j<1), \tag{19}$$

$$cn^{-(1/2)\rho(p+u)+1/2} \text{ when } M(x)=\exp(d|x|^i) \qquad (d\geqq 0,\ i<2). \tag{20}$$

A sufficient condition for uniform convergence on any given finite interval of the Hermite series when $M(x)=\exp(\frac{1}{4}x^2)$ for example, is that $p\geqq 2$, and a sufficient condition under the more restrictive hypothesis $M(x)=\exp(d|x|^i)$ $(i<2)$ is that $p\geqq 1$.

Consider now an extension, of the type mentioned in §4, of the results for the weight function $r(x)=\exp(-h|x|^\sigma)$. Let $r(x)=\exp(-h|x|^\sigma)$ be replaced by $r(x)=|\pi(x)|\exp(-h|x|^\sigma)$, $\pi(x)$ being any given polynomial. For this more general weight function, the expression (16) still gives an upper bound for the integral $I_{mn}$ and on any finite interval which contains no root of $\pi(x)$, the statements of all results on uniform convergence given for the weight function $\exp(-h|x|^\sigma)$ remain unaltered. Moreover, any interval

$(\alpha, \beta)$ on which $\pi(x)$ does vanish, may be enclosed in an interval $(a, b)$ on which $r(x)$ satisfies Hypothesis Y with $g=0$ and with $s$ equal to the greatest order of vanishing of $\pi(x)$ on $(\alpha, \beta)$. The upper bound for the error in the approximation given in the following corollary is obtained by use of the expression (10) of Theorem 2 because an upper bound for the error on $(a, b)$ is an upper bound for the error on the sub-interval $(\alpha, \beta)$ of $(a, b)$.

COROLLARY 3. *Suppose that* $r(x)=|\pi(x)|\exp(-h|x|^\sigma)$ $(h>0,\ \sigma>1)$ *where* $\pi(x)$ *is a polynomial having no root of order greater than* $s$ *on the given interval* $(\alpha, \beta)$ *and that* $M(x)=\exp(d|x|^i)$ $(d\geqq 0,\ i<\sigma)$. *Let* $\rho$ *be any assigned positive number less than* 1. *Then*

$$|f(x) - P_{mn}(x)| \leqq cn^{-\rho(p+u)(\sigma-1)/\sigma+(s+1)/m}$$

*uniformly on* $(\alpha, \beta)$.

The result of this corollary corresponds to the result of Corollary 1; a result corresponding to Theorem 4, which may be readily stated, is left for the reader.

**Remark.** When the weight function $r(x)$ tends toward zero very rapidly when $|x|$ becomes infinite, the effect on the magnitude of $I_{mn}$ of the interval $S=(-\infty, \infty)$ being infinite is less pronounced. For example, if $r(x)=\exp(-\exp x^2)$ and $M(x)=\exp(\exp|x|^i)$ $(i<2)$, then $A$ and $a$ may be chosen, in the steps illustrated by Lemmas 1 and 2 of this section, as constant multiples of $(\log n)^{1/2}$ and $(\log\log n)^{i/2}$ respectively so that $I_{mn}$ does not exceed $cn^{-m(p+u)/\omega}$ and hence, by Theorem 2, the error $|f(x)-P_{mn}(x)|$ does not exceed $cn^{-\rho(p+u)+1/m}$ where $\rho$ and $(1/\omega)$ are positive numbers less than 1. For the case of a finite interval $S$ and a weight function having a positive lower bound on $S$, the corresponding upper bounds for $I_{mn}$ and the error on any given interval interior to $S$ are $cn^{-m(p+u)}$ and $cn^{-\rho(p+u)+1/m}$ respectively. The details in the verification of the above statements are left for the reader.

6. **The simply infinite interval.** In this section, the minimized integral $I_{mn}$ is over a simply infinite interval $S$, taken for definiteness as $(0, \infty)$, and the functions $r(x)$, $M(x)$ and $f(x)$ satisfy the general hypotheses of §4 on this interval. In Lemma 1 below, an upper bound for $I_{mn}$ is obtained in terms of an integral, with an even weight function, over the doubly infinite interval in which $f(t^2)$ is the function to be approximated. The integral over the doubly infinite interval may then be bounded as is done in Theorem 5 below by use of Theorem 3.

LEMMA 1. *If* $p_n(t)$ *is any polynomial in* $t$ *of degree* $n$, *then*

$$I_{mn} \leqq 2\int_{-\infty}^{\infty} |t|\, r(t^2)\, |f(t^2) - p_n(t)|^m dt. \tag{21}$$

Let $\frac{1}{2}[p_n(t)+p_n(-t)]$, which is a polynomial of degree $n$ at most in $t^2=x$, be denoted by $q_n(x)$. Then $I_{mn}$ does not exceed the integral

$$\int_0^\infty r(x)\,|\,f(x)-q_n(x)\,|^m dx$$

$$=\int_{-\infty}^{\infty}|\,t\,|\,r(t^2)\,|\,\tfrac{1}{2}[f(t^2)-p_n(t)]+\tfrac{1}{2}[f(t^2)-p_n(-t)]\,|^m dt,$$

which in turn does not exceed the expression on the right in inequality (21). This establishes Lemma 1.

The following lemma is a result of Lemma 1 of §4 and of the formula for the $p$th derivative with respect to $t$ of $f(t^2)$. The details of the proof are omitted.

LEMMA 2. *The pth derivative with respect to t of $f(t^2)$ satisfies a Lipschitz condition of order u with coefficient $cM(a^2)a^{p+u}$ on every interval $-a\leqq t\leqq a$.*

Let $p_n(t)$ in Lemma 1 be the approximating polynomial for $f(t^2)$ corresponding to the even weight function $\rho(t)=|t|r(t^2)$ on the doubly infinite interval. Then by application of Theorem 3, $I_{mn}$ does not exceed $cH_{\mu\rho}(a, A)$ where $\mu(t)=M(t^2)\,|t|^{p+u}$. The expression (22) is not less than a constant multiple of the expression gotten by putting $x=t^2$ in $cH_{\mu\rho}(a, A)$ and then replacing $a^2$, $A^2$ by $a$, $A$ respectively; therefore $I_{mn}$ does not exceed a constant multiple of (22) as stated in the following theorem.

THEOREM 5. *If $S$ is the simply infinite interval $(0, \infty)$, the integral $I_{mn}$ does not exceed a constant multiple of*

$$[M(a)(aA)^{(p+u)/2}]^m\left[n^{-m(p+u)}+\int_A^\infty r(x)(4x/A)^{mn}dx\right]$$

$$+\int_a^\infty r(x)\,[M(x)x^{p+u}]^m dx. \tag{22}$$

The expression (22) corresponds to $H_{Mr}(a, A)$ of Theorem 3. The process of application of these results involves choosing $A$ and $a$ in terms of $n$ so that the integrals from $A$ to infinity and from $a$ to infinity respectively are negligible when $n$ becomes infinite. The non-negligible term of $H_{Mr}(a, A)$ is $[M(a)(A/n)^{p+u}]^m$ which is large with respect to the corresponding term $[M(a)(aA)^{(p+u)/2}/n]^m$ of (22), provided $a$ is small with respect to $A$. For example, let $r(x)=\exp(-h|x|^\sigma)$ and let $M(x)=\exp(d|x|^i)$ $(i<\sigma,\ h,\ d\geqq 0)$. Let $A$ and $a$ be chosen as $cn^{1/\sigma}$ and $c(\log n)^{i/\sigma}$ as in §5. Then if $S=(0, \infty)$,

$$|\,f(x)-P_{mn}(x)\,|\leqq cn^{-p(p+u)(\sigma-1/2)/\sigma+1/m} \tag{23}$$

uniformly on any given finite interval $(\alpha, \beta)$ where $\alpha$ and $\rho$ are positive and $\rho$ is less than 1. Uniform convergence results when the exponent of $n$ on the right in inequality (23) is negative and for this it is necessary that $\sigma$ exceed $\frac{1}{2}$ instead of 1 as was the case for the doubly infinite interval with the same $r(x)$ and $M(x)$. The result (23) corresponds to that of Corollary 1 of §5; comparison of these results shows first that for a given type of $r(x)$ and of $M(x)$, a higher order of uniform convergence is to be expected on the simply than on the doubly infinite interval and second that the types of admissible weight functions for the simply infinite interval include types of weight functions inadmissible on the doubly infinite interval.

Corresponding to each result of §5, there is a result for the simply infinite interval gotten by replacing $\sigma$ in §5 by $2\sigma$. When $r(x) = \exp(-x)$ on the interval $(0, \infty)$ and $m = 2$, for example, the results obtained apply to the Laguerre instead of the Hermite series. Theorems on the degree of uniform convergence on any finite interval including the point $x = 0$ may be developed by using expression (9) instead of expression (10) of Theorem 2 in the proofs. The details of these statements and the proofs are left for the reader.

7. **Essentially skew weight functions.** The integral $I_{mn}$ of this section is over the doubly infinite interval $S$. The discussion is intended to cover the case of essentially skew weight functions such as the exponential weight function discussed after the proof of Theorem 6 below; the discussion is, however, general and effects a unification of the results of §§4 to 6. The weight function $r(x)$ is subject only to general conditions of integrability such as those of §4 but $r(x)$ is not in general an even function of $x$. The function $f(x)$ is assumed throughout to have a $p$th derivative $f^{(p)}(x)$ which satisfies a Lipschitz condition of order $u$ with coefficient $M(a)$ on $(0, a)$ and with coefficient $M(-a)$ on $(-a, 0)$ for every $a$, where $M(x)$ and $M(-x)$ are non-negative and non-decreasing for non-negative values of $x$. It is assumed as usual that $f(x)$ and its first $p$ derivatives vanish when $x = 0$.

Let it be explicitly understood that $B$ is taken as zero, in Lemma 1 or in Theorem 6 of this section, only when $r(x)$ is identically zero for negative values of $x$ and that the integral over $(-\infty, -B)$ in (25) is then zero. With this understanding, if $B = 0$, the upper bound for $I_{mn}$ of Theorem 6 reduces essentially to the upper bound of Theorem 5 for the minimized integral over the simply infinite interval. On the other hand, when $r(x)$ and $M(x)$ are even functions of $x$ and $a = b$ and $A = B$, then Theorem 6 gives essentially the upper bound of Theorem 3 for the minimized integral over the doubly infinite interval with an even weight function. Thus, Theorems 3 and 5 are essentially special cases of the more general Theorem 6.

A skew exponential weight function is used to illustrate the application of Theorem 6; the application follows the proof of the theorem which has for its basis the following lemma concerning the approximation to a particular type of function $F(x)$.

LEMMA 1. *Suppose that $F(x)$ has a pth derivative which satisfies a Lipschitz condition of order $u$ with coefficient $L$ for all real values of $x$ and that $|F(x)|$ nowhere exceeds $\mu$ and is identically zero except perhaps on $(-l_1, l_2)$. Then there exists a polynomial $Q_n(x)$ of degree $n>p$ such that*

$$\int_{-\infty}^{\infty} r(x)\,|F(x)-Q_n(x)|^m dx \leqq c\left\{L^m[A(l_2+B)/n^2]^{m(p+u)/2} + \mu^m\left[\int_{-\infty}^{-B} r(x)(-4x/B)^{mn}dx + \int_{A}^{\infty} r(x)(4x/A)^{mn}dx\right]\right\}$$

*whenever $0\leqq l_1\leqq B$, $0<l_2\leqq A$ and $B\leqq A$.*

Let $y=x+B=t^2$ and let $F(x)\equiv\phi(y)\equiv\psi(t)$. Then by Lemma 2 of §6, $\psi^{(p)}(t)$ satisfies a Lipschitz condition of order $u$ with coefficient $L(B+l_2)^{(p+u)/2}$ on $-\infty<t<\infty$ so that by Theorem J, there exists a polynomial $p_n(t)$ of degree $n$ such that on $-(A+B)^{1/2}\leqq t\leqq(A+B)^{1/2}$,

$$|\psi(t)-p_n(t)|\leqq cL[(B+l_2)(A+B)/n^2]^{(p+u)/2}. \tag{24}$$

Take* the polynomial $Q_n(x)$ as

$$Q_n(x)\equiv q_n(y)\equiv q_n(t^2)\equiv\tfrac{1}{2}[p_n(t)+p_n(-t)].$$

Since $\psi(t)$ is an even function of $t$, the error $F(x)-Q_n(x)$ on $(-B, A)$ which is equal to $\psi(t)-q_n(t^2)$, does not exceed in magnitude the expression on the right in (24) which in turn does not exceed $cL[A(l_2+B)/n^2]^{(p+u)/2}$. It may be assumed as usual that the magnitude of $Q_n(x)$ does not exceed $2\mu$ on $(-B, A)$ and therefore does not exceed $2\mu(4x/A)^n$ and $2\mu(-4x/B)^n$ on $(A, \infty)$ and $(-\infty, -B)$ respectively. Integration now yields the inequality of the lemma.

THEOREM 6. *Under the hypotheses of this section on $r(x)$ and $f(x)$, the minimized integral $I_{mn}$ over the doubly infinite interval does not exceed a constant multiple of*

$$[(M(a)+M(-b))(A(a+B))^{(p+u)/2}]^m\left\{n^{-m(p+u)} + \int_{-\infty}^{-B} r(x)(-4x/B)^{mn}dx + \int_{A}^{\infty} r(x)(4x/A)^{mn}dx\right\} \tag{25}$$

* This device was suggested by Professor J. V. Uspensky.

$$+ \int_{-\infty}^{-b} r(x)[M(x)(-x)^{p+u}]^m dx + \int_a^{\infty} r(x)[M(x)x^{p+u}]^m dx$$

$$+ [(M(a) + M(-b))(a+b)^{p+u}]^m \int_a^{2a} r(x)dx$$

*wherever* $0 \leqq b \leqq B$, $2 \leqq 2a \leqq A$ *and* $B \leqq A$.

The minimized integral $I_{mn}$ does not exceed

$$2^m \int_{-\infty}^{\infty} r(x) \left| f(x) - F(x) \right|^m dx + 2^m \int_{-\infty}^{\infty} r(x) \left| F(x) - Q_n(x) \right|^m dx \tag{26}$$

where $F(x)$ and the polynomial $Q_n(x)$ of degree $n$ are selected arbitrarily. Let $F(x)$ be defined as the function which is zero when $x$ is not on $(-b, 2a)$, as $f(x) - V_p(x, -b)$ on $(-b, a)$, and on $(a, 2a)$ as the polynomial of degree $(2p+1)$ which agrees with $F(x)$ and its first $p$ derivatives at $x=a$ and has a root of order $(p+1)$ at $x=2a$. Then $F(x)$ satisfies the conditions of Lemma 1 with $L=c[M(-b)+M(a)]$, $l_1=b$, $l_2=2a$, and since $a \geqq 1$, $\mu = cL(a+b)^{p+u}$. Let $Q_n(x)$ be chosen as in Lemma 1. Then the second integral in (26) does not exceed a constant multiple of the first term in the expression (25). The first integral in (26) does not exceed a constant multiple of the remaining three terms in (25). Therefore $I_{mn}$ does not exceed the expression (25). This completes the proof of Theorem 6.

To illustrate the application of Theorem 6, consider the exponential weight function $r(x)$ equal to $\exp(-h|x|^{\sigma})$ and $\exp(-h'|x|^{\sigma'})$ $(h, h' > 0;$ $\sigma, \sigma' > \frac{1}{2})$ for negative and non-negative values of $x$ respectively. Let $M(x)$ be equal to $\exp(d|x|^{i})$ and $\exp(d'|x|^{i'})$ $(d, d' > 0;\ i < \sigma;\ i' < \sigma')$ for negative and non-negative values of $x$ respectively. When $A = cn^{1/\sigma}$, $B = cn^{1/\sigma'}$, $a = c(\log n)^{i/\sigma}$ and $b = c(\log n)^{i'/\sigma'}$, with the $c$'s suitably chosen, then it is found by use of Theorems 6 and 2 that if $\rho$ is any assigned positive number less than 1 and if $(\alpha, \beta)$ is a given finite interval, then

$$\left| f(x) - P_{mn}(x) \right| \leqq cn^{-(1/2)\rho[2-(1/\sigma+1/\sigma')]+1/m} \tag{27}$$

uniformly on $(\alpha, \beta)$. This result corresponds to that of Corollary 1 §5; a result corresponding to Theorem 4 may be stated; this statement as well as the details of the verification of the statements made above are left for the reader.

In conclusion, notice that the admissibility of the above skew weight function requires only that $(1/\sigma + 1/\sigma') < 2$ so that in the special case of an even weight function, $\sigma = \sigma'$, it is necessary that $\sigma = \sigma'$ exceed 1 as was the case in §5. If $\sigma$ and $\sigma'$ both exceed 1 and $\sigma' < \sigma$ then the weight function is

admissible on the basis of §5 but $(1/\sigma+1/\sigma')$ in (27) must be replaced by $(2/\sigma')$ which increases the upper bound for the error essentially. If $\sigma'$ is between $\frac{1}{2}$ and 1, then the weight function is inadmissible on the basis of §5 but is admissible here if $\sigma$ exceeds $\sigma'/(2\sigma'-1)$; the restriction on $r(x)$ for large positive values of $x$ can thus be weakened at the expense of increasing the restriction on $r(x)$ for large negative values of $x$ and vice versa.

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

# ELEMENTARY PROPERTIES OF THE $t_\nu$-FUNCTIONS*

BY

W. E. MAIER†

The uniformization of the elliptic integral of the first species is given in classical form by the following expression of Weierstrass:

$$\lim_{n\to\infty} \sum_{h,k}^{h^2+k^2<n} (u + \omega_1 h + \omega_2 k)^{-2} = \wp\left(u, \frac{\omega_1}{2}, \frac{\omega_2}{2}\right).$$

This leads to building up rationally an initial value of this $\wp$-function from the coefficients of the given biquadratic polynomial. The projective invariants of biquadratic forms and their covariants are connected through identical relations which may be regarded as characteristic functional equations for elliptic functions. However, this method does not seem to reduce matters to their elements.

Making use of some earlier investigations‡ we shall link up the above uniformization with certain double series of the following form:

$$\lim_{n\to\infty} \sum_{h,k}^{0<h^2+k^2<n} \frac{e^{2\pi i(xh+yk)}}{(h\omega_1 + k\omega_2)^\nu} = t_\nu(x, y) \qquad (\nu = 1, 2, \cdots).$$

By means of this expression we are able to formulate the known properties of elliptic functions in a more natural way. In particular we shall derive the algebraic relation between $t_1(x, y)$ and $t_1(x/2, y/2)$ which is the simplest case of a division theory, the latter containing the corresponding developments of Abel.

Moreover, these functions seem to be of interest for their own sake because of the existence of relations of the following type:

$$t_4\left\{x + \frac{1}{\pi i} t_1\left(\frac{x}{2}, \frac{y}{2}\right),\ y + \frac{\omega}{\pi i} t_1\left(\frac{x}{2}, \frac{y}{2}\right)\right\} = 0,$$

which for $\omega = \rho^{2\pi i/3}$ holds identically in $x, y$.

Let $w$ be complex, $\nu = 0, 1, 2, 3, 4$, and $\tau$, complex. We form the polynomial biquadratic in $w$

---

* Presented to the Society, April 18, 1930; received by the editors March 12, 1930.

† Research Fellow of the International Education Board.

‡ Mathematische Zeitschrift, 1930, *Zur Theorie der elliptischen Funktionen.*

$$\sum_{\nu=0}^{4} \frac{w^\nu}{\nu!}\tau_{4-\nu} = V(w), \tag{1}$$

and the expressions

$$24\tau_0\tau_4 - 24\tau_1\tau_3 + 12\tau_2^2 = (24)^2 \cdot S, \tag{2}$$

$$48\tau_0\tau_2\tau_4 + 24\tau_1\tau_2\tau_3 - 8\tau_2^3 - 36\tau_0\tau_3^2 - 24\tau_1^2\tau_4 = (24)^3 \cdot T, \tag{2'}$$

which are invariant under linear fractional transformations of $w$.

Now the function $V^{1/2}$ which is two-valued in $w$ is to be made one-valued in its dependence on a complex variable $u$ by means of elliptic functions. The latter can be introduced in connection with two complex auxiliary magnitudes $\omega_1$, $\omega_2$ restricted by $F(\omega_2/\omega_1) > 0$. Using the notation

$$\sum_{h,k}^{0<h^2+k^2} = \sum_{h,k}'$$

we write a system of two transcendental determining equations

$$\left(\frac{24}{\tau_0}\right)^2 \frac{S}{60} = \sum'(\omega_1 h + \omega_2 k)^{-4},$$
$$\left(\frac{24}{\tau_0}\right)^3 \frac{T}{140} = \sum'(\omega_1 h + \omega_2 k)^{-6}, \tag{3}$$

and we construct in the customary way the Weierstrassian expression

$$\lim_{n\to\infty} \sum_{h,k}^{h^2+k^2<n} (u + \omega_1 h + \omega_2 k)^{-2} = \wp\left(u, \frac{\omega_1}{2}, \frac{\omega_2}{2}\right) = \wp(u).$$

The complex number $z$ is then found as the unique solution mod $\omega_1$, $\omega_2$ of the additional determining equation

$$\tau_1^2 - 2\tau_0\tau_2 = \tau_0^2 \cdot \wp(z), \tag{4}$$

and it also appears that

$$\tau_0^3 \wp'(z) = 2\tau_1^3 - 6\tau_0\tau_1\tau_2 + 6\tau_1^2\tau_3. \tag{4'}$$

The above uniformization is then expressed by

$$w = -\frac{\tau_1}{\tau_0} + \frac{1}{2}\,\frac{\wp'(u) - \wp'(z)}{\wp(u) - \wp(z)}, \tag{5}$$

$$V^{1/2} = \tau_0^{1/2}[\wp(u) - \wp(u+z)]. \tag{5'}$$

Between the two statements (5), (5') there exists moreover the connection

$$\frac{1}{2}\frac{d}{du}\frac{\wp'(u)-\wp'(z)}{\wp(u)-\wp(z)}=\wp(u)-\wp(u+z), \tag{6}$$

which, in consequence of the addition theorem, leads to the integral representation

$$\tau_0^{1/2}\int\frac{dw}{V^{1/2}}=u(w)=u.$$

This whole situation suggests a further problem concerning the number of the undetermined quantities involved. Without loss in generality we may place $\tau_0=-1$; then through a unimodular linear transformation the quadruple of coefficients $\tau_1, \cdots, \tau_4$ in (1) is simultaneously transformed into another system $\tau_1^*, \cdots, \tau_4^*$; in this way the quantities on the right in (2), (2′) remain unchanged. Now in making, according to the classical theory, the transition from the old to the new quadruple we have two degrees of freedom. These we wish to put in evidence. We succeed in doing so by introducing four new parameters $x, y$; $\Omega_1, \Omega_2$ through the following conditions:

(I) There is a one-to-one correspondence between $\tau_1, \cdots, \tau_4$ and $x, y$; $\Omega_1, \Omega_2$.

$$\text{(II)}\qquad \frac{\partial S}{\partial x}=\frac{\partial S}{\partial y}=\frac{\partial T}{\partial x}=\frac{\partial T}{\partial y}=0.$$

In other words, we wish to map the $\tau$-quadruple on two number pairs in such a way that the statements (2), (2′) hold identically in one of them. This plan we carry out by making use of the periods implied by (3) placing $\Omega_j=\omega_j$ for $j=1, 2$. Furthermore let $0<x<1$, $0<y<1$;

$$\lim_{n\to\infty}\sum_{h,k}^{0<h^2+k^2<n}=\sum' \qquad (\nu=1, 2, 3, \cdots);$$

$$\sum'\frac{e^{2\pi i(xh+yk)}}{(\omega_1 h+\omega_2 k)^\nu}=t_\nu(x, y)=t_\nu. \tag{7}$$

In particular for $\nu=4$ there exists

$$\lim_{x,y\to 0} t_4=t_4(0, 0). \tag{8}$$

Also, in addition to the previous condition $\tau_0=-1$ we place $\tau_\lambda=t_\lambda$ for $\lambda=1, 2, 3$ and $\tau_4=t_4-t_4(0, 0)$. As analogous to (7), we still use the modular form $t_6(0, 0)$, which is of the $-6$th dimension. The functions $t_\nu$ introduced by (7) occupy an intermediate position between modular forms and elliptic functions. The latter can be represented rationally through the $t_\nu$, but not conversely. The expressions (7) lead to a theory which goes beyond the results

of Weierstrass. They represent meromorphic functions in $x$ and $y$ which are successively joined by the differential relation

$$\left\{\frac{\omega_1}{2\pi i}\frac{\partial}{\partial x}+\frac{\omega_2}{2\pi i}\frac{\partial}{\partial y}\right\}t_\nu = t_{\nu-1}, \tag{9}$$

which for a given $t_{\nu-1}$ admits of the determination of $t_\nu$ but for an additive term. The latter is not only meromorphic in each of the variables $x$, $y$ but is even an analytic function of the single argument $(\omega_1 y-\omega_2 x)$. If we add as a condition in the large (Bedingung im Grossen) the periodicity

$$t_\nu(x, y) = t_\nu(x+1, y) = t_\nu(x, y+1), \tag{10}$$

then this meromorphic term is found to be twofold periodic, that is, elliptic, and hence to be characterized through limiting conditions. The initial value (8), together with relation (9), the periods (10), and e.g. the limiting condition

$$\lim_{x,y\to 0}\left\{\frac{1}{\omega_1 y-\omega_2 x}-t_1\right\} = 0, \tag{10'}$$

is sufficient to determine $t_1(x, y)$.

The above $t_\kappa$ are not mutually independent but are connected by algebraic recursions. Among others we have

$$9t_3^2+6t_3[t_1^3+3t_1t_2]-\{8t_2^3+3t_1^2t_2^2 - 15t_4(0,0)[t_1^2+2t_2]+35t_6(0,0)\} = 0, \tag{11}$$

and

$$t_2^2-2t_1t_3-2t_4-3t_4(0,0) = 0, \tag{11'}$$

which allow us to transform the invariants (2),(2′) of the polynomial (1). According to (7), (8) the individual summands on the left side of (2),(2′) depend on $x$, $y$; their sums, however, as seen from (3), are independent of $x$, $y$, a fact which we recognize as equivalent to (11), (11′). For

$$-24[t_4-t_4(0,0)]-24t_1t_3+12t_2^2 = 12[t_2^2-2t_1t_3-2t_4+2t_4(0,0)] = 60t_4(0,0),$$

$$\begin{aligned}&-48t_2[t_4-t_4(0,0)]+24t_1t_2t_3-8t_2^3+36t_3^2+24t_1^2[t_4(0,0)-t_4]\\&=12[t_1^2+2t_2]\{5t_4(0,0)+2t_1t_3-t_2^2\}+24t_1t_2t_3-8t_2^3+36t_3^2\\&=-140t_6(0,0).\end{aligned}$$

Moreover from (4), (4′), (7), (8) we infer

$$\begin{aligned}t_1^2+2t_2 &= \wp(z),\\ 2t_1^3+6t_1t_2+6t_3+\wp'(z) &= 0,\end{aligned} \tag{12}$$

and from (10), (10′) obtain the value of the argument in (4),

$$z = \omega_1 y - \omega_2 x. \tag{13}$$

Substituting this value into the following identities:

$$t_1 + \frac{1}{2}\frac{\wp'(u) - \wp'(z)}{\wp(u) - \wp(z)} = w; \quad \wp(u) - \wp(u+z) = -V^{1/2}; \tag{14}$$

$$\sum_{\nu=0}^{4} \frac{w^\nu}{\nu!} t_{4-\nu} = t_4(0, 0) + V(w), \tag{15}$$

we find the addition theorem of the $\wp$-function translated into the $t_\nu$-language.

If the $\tau_\nu$ quadruple is given through (1), the invariants may be computed from (11),(11′); and $\wp(z)$ from (12); and thus a suitable triplet $\omega_1, \omega_2, z$ can be determined as usual. The real values $x, y$ then follow uniquely from (13).

If $\lambda = 1, 2, \cdots$, for a sufficiently small $|w|$ and for $\omega_1 y - \omega_2 x \not\equiv 0 \pmod{\omega_1, \omega_2}$ Taylor's theorem in two variables furnishes, as an immediate consequence of the differential relation of the original series,

$$\begin{aligned} t_\lambda\left(x + w\frac{\omega_1}{2\pi i},\ y + w\frac{\omega_2}{2\pi i}\right) &= \sum_{\nu=0}^{\infty} \frac{w^\nu}{\nu!}\left\{\frac{\omega_1}{2\pi i}\frac{\partial}{\partial x} + \frac{\omega_2}{2\pi i}\frac{\partial}{\partial y}\right\}^\nu t_\lambda(x, y) \\ &= \sum_{\nu=0}^{\infty} \frac{w^\nu}{\nu!} t_{\lambda-\nu}(x, y) \\ &= \sum_{\nu=0}^{\lambda} \frac{w^\nu}{\nu!} t_{\lambda-\nu}(x, y). \end{aligned} \tag{16}$$

On the other hand if we develop a power series with indeterminate coefficients we obtain algebraic recursions between any three $t_\lambda$ so that the identity (16) yields an independent theory of the $t_\lambda$ functions.

The special case $\lambda = 4$ gives

$$V(w) = t_4\left(x + w\frac{\omega_1}{2\pi i},\ y + w\frac{\omega_2}{2\pi i}\right) - t_4(0, 0). \tag{17}$$

Studying the analytic continuation* of this function into the two complex $y$ and $x$-planes we find that the right side of (17) is a function of $z$ only, and even elliptic in $z$. By putting

$$u = \omega_1 r - \omega_2 q \tag{13′}$$

* See Mathematische Zeitschrift, loc. cit.

which means a splitting of $u$ according to periods, we can give a linear representation

$$w = t_1(q + x, r + y) - t_1(q, r) \tag{14'}$$

where $w$ is itself an argument of the similar function (17).

We next wish to emphasize the fact which characterizes the elliptic case, that two projective independent invariants of $V$ remain invariant also under continuous $x$, $y$ transformations. For this purpose we introduce further the covariants of $V$ and the algebraic relation connecting them. In forming successive derivatives with respect to $w$, we are lead to the Hessian function which, as $V$, is elliptic both in $u$ and $z$. Moreover this covariant $H$ has two further properties in common with the original polynomial: The operation (9) transforms both times each coefficient into its right neighbor; to the final coefficient $\tau_0 = -1$ in (1) corresponds $2t_1^2 + 4t_2$ in (16'). Starting from the two simplest elliptic functions 1 and $\wp'$ we thus reach, after four integrations with suitable determination of the constants, the functions $V$ and $H$ respectively. A similar situation prevails for the Jacobian; the degree of its coefficients is increased by 3 in $t_1$, the order of the form by 2 in $w$.

The determination of these constants we shall not carry out here; but we nevertheless note that, taking the first statement in (16'), we get

$$H = 3t_3^2\left(x + w\frac{\omega_1}{2\pi i},\ y + w\frac{\omega_2}{2\pi i}\right) + 4\left[t_4(0,0) - t_4\left(x + w\frac{\omega_1}{2\pi i},\ y + w\frac{\omega_2}{2\pi i}\right)\right]t_2\left(x + w\frac{\omega_1}{2\pi i},\ y + w\frac{\omega_2}{2\pi i}\right)$$

where the $t_2$, $t_4$ may be eliminated by means of the recursion

$$5[t_6 - t_6(0,0)] + 3t_2[t_4 - t_4(0,0)] = 2t_3^2.$$

Now let us consider the biquadratic equation

$$V(w) = 0, \tag{17'}$$

which determines discrete values of $w$. The corresponding determining equation for $u$ is, according to (14),

$$\wp(u) = \wp(u + z).$$

The arguments in this equation are connected by the relations

$$-u \equiv u + z, \quad u \equiv -\frac{z}{2} \pmod{\omega_1, \omega_2}.$$

On the other hand (17′) will be satisfied, according to (10), (14′), by

$$w = t_1\left(\frac{x}{2}, \frac{y}{2}\right) - t_1\left(-\frac{x}{2}, -\frac{y}{2}\right) = 2t_1\left(\frac{x}{2}, \frac{y}{2},\right) \tag{18}$$

and the pairs of arguments

$$\left(\frac{x+1}{2}, \frac{y}{2}\right); \quad \left(\frac{x}{2}, \frac{y+1}{2}\right); \quad \left(\frac{x+1}{2}, \frac{y+1}{2}\right)$$

together with (18) furnish the complete system of solutions to (17′).

This is the simplest example of the division theory of the functions $t_\nu$. It corresponds to Abel's theory of the division of arguments for elliptic functions. As in the case of the $t_1$ function the bisection is accomplished through an equation of the fourth degree; but in the present theory the structure of the coefficients used is more obvious.

The limiting case $|\omega_2| \to \infty$ admits of a simple check on (17′), (18); the expression (7) becomes a simply infinite series whose sum is a polynomial in $x$. In agreement with (1), (17′), (18) we have in particular,

$$\omega_1 w = 2\pi i(1 - x); \quad \omega_1\tau_1 = \pi i(1 - 2x),$$

$$\omega_1^2\tau_2 = \frac{\pi^2}{3}(1 - 6x + 6x^2),$$

$$\omega_1^3\tau_3 = \frac{2\pi^3 i}{3}(x - 3x^2 + 2x^3),$$

$$\omega_1^4\tau_4 = -\frac{2\pi^4}{3}(x^2 - 2x^3 + x^4).$$

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

# ON DIVISION ALGEBRAS*

BY

O. C. HAZLETT

1. **Summary.** Linear algebras have been studied in great detail for many years, but they have usually been studied over a field $F$. In the papers by Dickson,† Cecioni,‡ A. A. Albert§ and others, a division algebra over a field $F$ can be regarded as a division algebra defined over an algebra containing $F$. In the present paper, it is proved that every division algebra over a field $F$ can be regarded as an algebra $E_1$ over a division algebra $D_1$, where $D_1$ is in its turn an algebra $E_2$ over an algebra $D_2$, etc. where each of the component algebras $D_i$ is a division algebra defined over a field $F_i$ contained in $E_i$ of one of the three following types:

(1) fields,

(2) normal Dickson algebras,

(3) algebras which contain no Dickson subalgebras.

This paper does not determine whether there actually exist algebras of the third type, but the writer has the feeling that none such exist. This principle is applied to certain solvable algebras and it appears that under suitable conditions a solvable algebra is a direct product of solvable algebras.

2. **Relation to the literature; definitions.** We shall use the term *field* to denote a set of elements with two operations called addition and multiplication which is an abelian group with respect to addition and which, when we exclude the zero element (=the unit as to addition), is an abelian group with respect to multiplication. If, however, we do not insist on multiplication being commutative, then a set of numbers which is closed under the four rational operations (excluding division by zero) is said to form a *division (or primitive) algebra*, if it be always possible to solve for $x$ in the algebra

* Presented to the Society, May 2, 1925; received by the editors in April, 1930.

† *Linear associative algebras and abelian equations*, these Transactions, vol. 15 (1914), pp. 31–46; *New division algebras*, ibid., vol. 28 (1926), pp. 207–234; *New division algebras*, Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 555–560.

‡ *Sopra un tipo di algebre prive di divisori dello zero*, Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 209–254.

§ *On the structure of normal division algebras*, Annals of Mathematics, (2), vol. 30 (1929), pp. 322–338; *Normal division algebras in* 4 $p^2$ *units, p an odd prime*, ibid., pp. 583–590.

both of the equations $xy=z$ and $yx=z$ when $y$ and $z$ are any two numbers in the algebra. These are division algebras over any field $F$.*

When the field $F$ is the field $R$ of reals, it is well known that the only division algebras are (1) an algebra of one unit, the field $R$ itself; (2) an algebra of two units, 1 and $i$, the field $C$ of ordinary complex numbers; (3) an algebra of four units $(1, i, j, k)$, real quaternions. When the field of definition is $C$, the only division algebra is $C$ itself. But when $F$ is any other field, the nature and number of types of division algebras has not yet been fully determined, although recently much work has been done.†

When $F$ is any algebraic field, Dickson considered a general class of algebras that he called Type $A$. He defines an algebra of this type as any linear associative algebra $A$, the coördinates of whose numbers range over $F$ and for which the following properties hold:

(a) There exists in $A$ a number $i$ satisfying an equation $\phi(x)=0$ of degree $n$ with coefficients in $F$ and irreducible in $F$.

(b) Any number of $A$ which is commutative with $i$ is in $F(i)$.

(c) There exists in $A$ a number $j$, not in $F(i)$, such that $ji=\theta j+\sigma$, where $\theta$ and $\sigma$ are in $F(i)$.

All three of these conditions are satisfied by real quaternions, and the first two by any linear associative division algebra $D$ over $F$, where we take $i$ so that the degree of the irreducible equation in $F$ satisfied by $i$ is a maximum. Dickson‡ showed that every algebra of type $A$ over $F$ has a subalgebra $S$ over $F$ which can be exhibited as an algebra $L$ over a field $K$ with units $i^s j^k (k, s=0, \cdots, r-1)$. Multiplication is defined by the relations $ji=\theta(i)j$, $j^r=g$ where $i$ is an element of $S$ satisfying in the field $K$ a uniserial abelian equation of degree $r$, with the roots $i, \theta(i), \theta^2(i), \cdots, \theta^{r-1}(i)$, where $\theta^r(i)=i$ and where $g$ is a number of $K$. Dickson showed for $r=2, 3$ that $g$ could be so chosen that in the algebra $L$, division (except by zero) is always possible and unique. Wedderburn,§ a few months later, showed that, for

* Wedderburn, *On hypercomplex numbers*, Proceedings of the London Mathematical Society, (2), vol. 6 (1907), p. 91; *A type of primitive algebra*, these Transactions, vol. 15 (1914), pp. 162–166; *On division algebras*, ibid., vol. 22 (1921), pp. 129–135; Dickson, *Linear Algebras* (Cambridge Tract No. 16); *Linear associative algebras and abelian equations*, these Transactions, vol. 15 (1914), pp. 31–46; *Algebras and their Arithmetics*, chapter V, Appendices I, II; *Algebren und ihre Zahlentheorie*, Zürich, Füssli, 1927, chapter III, etc.; F. Cecioni, *Sopra un tipo di algebre prive di divisori dello zero*, Rendiconti del Circolo Matematico di Palermo, vol. 47 (1923), pp. 209–254; Hazlett, *On the theory of associative division algebras*, these Transactions, vol. 18 (1917), pp. 167–176.

† In addition to the papers cited above, see E. Artin, *Über einen Satz von Herrn J. H. Maclagan Wedderburn*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 5 (1927), pp. 245–250; *Zur Theorie der hyperkomplexen Zahlen*, ibid., pp. 251–260; R. Brauer, *Über Systeme hyperkomplexer Zahlen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 79–107.

‡ These Transactions, vol. 15 (1914), p. 37.

§ These Transactions, vol. 15 (1914), pp. 162–166.

general $r$, $g$ can be so chosen that $L$ (and hence $S$) is a division algebra. Wedderburn has called division algebras of the type of $L$ *Dickson algebras.**

A *quadrate* (*normal*) *division algebra* is any division algebra which contains no numbers other than scalars which are commutative with every number of the algebra. Any division algebra is the direct product of a field and a quadrate division algebra.†

Several years later, Wedderburn‡ proved that every division algebra of order 9 is either a field or a Dickson algebra. In a footnote of this paper he says that, although he has been unable to construct a division algebra analogous to Dickson algebras, but corresponding to a non-cyclic abelian equation, yet "it appears probable that they exist." More recently, Cecioni§ considered division algebras over any field $K$ whose order is a perfect square, $n^2$, and which contain an element $i$ whose reduced equation is an abelian equation of degree $n$ with coefficients in $K$, and determined certain properties of them. In April 1926, Dickson published another beautiful paper‖ on division algebras, in which he proved that just as there is a Dickson algebra corresponding to any cyclic group, so there is a division algebra corresponding to any solvable group.

More recently, A. A. Albert¶ considered normal division algebras of order $n^2$ over a non-modular field $F$ of type $R_k$.** He proved (Theorem 21): Let $A=B\times C$, where $B$ is an associative normal division algebra of order $n^2$ over $F$, of type $R_n$, and $C$ is an associative normal division algebra of order $m^2$ over $F$, of type $R_m$. If $A$ be an associative division algebra, then $A$ is a normal division algebra of type $R_{mn}$.††

---

* These Transactions, vol. 22 (1921), p. 134.

† Wedderburn, these Transactions, vol. 22 (1921), p. 130.

‡ *On division algebras*, these Transactions, vol. 22 (1921), pp. 129–135.

§ Loc. cit.

‖ *New division algebras*, these Transactions, vol. 28 (1926), pp. 207–234.

¶ Loc. cit. I, p. 322.

** A normal division algebra is said to be of *type* $R_k$ if it contain an element $x$ of grade $n$ and type $R_k$. The element $x$ is said to be of *grade* $t$ if the degree of its minimum equation is $t$; and of *type* $R_k$ if its minimum equation have a set of $k$ ordinary complex roots which are polynomials in one of the set.

†† In his proof that $A$ is normal, he uses the assumption that $B$ and $C$ are of special types. If we define a normal algebra as any algebra $A$ (not necessarily a division algebra) over $F$ which is such that the only numbers commutative with every number of $A$ are in $F$, then we easily prove

THEOREM 1. *If $B$ and $C$ be two normal algebras defined over the field $F$, then $A=B\times C$ is normal*

For let $a=\sum b_ic_i$ be a number of $A$ commutative with every number of $A$ where each $b_i<B$ and each $c_i<C$. We may assume without loss of generality that the $b_i$ are linearly independent with respect to $F$. Since $a$ is commutative with every number $\gamma<C$, then $\sum b_i(c_i\gamma)=\sum b_i(\gamma c_i)$. By the definition of a direct product, the $b_i$ are linearly independent with respect to $C$ and thus $c_i\gamma=\gamma c_i$. Since $C$ is normal, $c_i<F$ and $a<B$. Since $a$ is commutative with every number of $B$, then $a<F$ and

In the second paper,* he shows that every normal division algebra of order $4p^2$, $p$ an odd prime, of the special type considered by Dickson† is of type $R_{2p}$ and hence is "known," as he calls it. A "known" algebra might be described as a division algebra $\Gamma$ which is a Dickson algebra with respect to a proper subalgebra $\Sigma$ which, in its turn, is a "known" algebra with respect to a proper subalgebra.‡

---

$A$ is normal.

Similarly in his proof that $A$ is of type $R_{nm}$, he uses the assumption that $F(i)$ and $F(k)$ are imbedded in division algebras of a special kind, but this is not necessary. In fact, it is easy to prove

THEOREM 2. *Let $F(i)$ be a field of order $n$ defined by $i$ of type $R_n$ and let $F(k)$ be a field of order $m$ defined by $k$ of type $R_m$. Then any number $t$ that defines $F(i, k)$, of order $p$, is of type $R_p$.*

For $F(i, k)$ is defined as an overfield of $F$ by any number $t$ satisfying an irreducible equation of degree $p$, say

$$t=t_{11}=\sum a_\beta(i)k^\beta \qquad (\beta=0, \cdots, m-1)$$

where $a_\beta(i)<F(i)$. Then

$$t_{rs}=\sum a_\beta(\theta_r)\phi_s{}^\beta \qquad (r=1, \cdots, n;\ s=1, \cdots, m),$$

where $\theta_1, \cdots, \theta_n$ is a complete set of conjugates of $i$ with respect to $F$ and $\phi_1, \cdots, \phi_m$ is a complete set of conjugates of $k$ with respect to $F$, is in $F(i, k)$ and thus is a polynomial in $t$ with coefficients in $F$. Hence $t$ is a root of $\phi(\xi)=\pi(\xi-t_{rs})=0$ of degree $mn$ and with coefficients in $F$. Since the minimum function of $t$ is a factor of $\phi(\xi)$, then $t$ is of type $p$.

* Loc. cit. II, p. 583.

† *New division algebras*, Bulletin of the American Mathematical Society, vol. 34, p. 555.

‡ Several of his theorems assert that if $x$ be a number of a normal division algebra $A$ satisfying certain conditions, then $A$ contains a number satisfying certain other conditions and the proof of each of these theorems (32, 34, 35) actually uses numbers of $A$ outside of $F(x)$, although the theorem seems to be about $F(x)$ rather than about $A$. Theorem 32: *Let $A$ be normal division algebra. Let $x<A$ have the property that $-x$ satisfies the minimum equation of $x$. Then the minimum equation of $x$ has even powers only. Let $x^2$ be of grade $p$ with $\phi(\omega)=0$ as its minimum equation. Then $x$ has grade $2p$ and the minimum equation is $\phi(\omega^2)=0$.* We can readily prove

THEOREM 3. *Let $x$ generate a field $F(x)$ and let $x$ have the property that $-x$ satisfies the minimum equation of $x$. Then the minimum equation of $x$ has even powers only; and if $x^2$ be of grade $p$ with $\phi(\omega)=0$ as its minimum equation, then $x$ has grade $2p$ and its minimum equation is $\phi(\omega^2)=0$.*

For if $\chi(\xi)=0$ be the minimum equation of $x$, then $-x$ is a root of $\chi(\xi)=0$ and thus $x$ is a root of $\chi(-\xi)=0$. Since $\chi(-\xi)=0$ is of the same degree as $\chi(\xi)=0$, the minimum equation, they must be identical except for a change of sign throughout, and thus $\chi(\xi)=0$ contains only even powers of $\xi$. By hypothesis, $F(x^2)$ is of order $p$ and, since $x$ satisfies an equation of degree at most 2 with respect to $F(x^2)$, then $F(x)$ is of order $p$ or $2p$. But, if the order were $p$, then $x$ would equal a polynomial in $x^2$ of degree $p$. Transposing the term $x$, we have an equation $\psi(\xi)=0$ containing precisely one term of odd degree with non-zero coefficient. Since $\chi(\xi)$ is a factor of $\psi(\xi)$, then $-x$ must satisfy $\psi(\xi)=0$ which implies $x=-x$ and this is impossible. Hence $F(x)$ is of order $2p$ and thus its minimum equation is $\phi(\omega^2)=0$. Instead of his Theorem 34, we prove the stronger

THEOREM 4. *Let $F(x)$ be a field of order $2p$ and let $x$ satisfy*

$$\phi(\omega)\equiv\omega^{2p}+\alpha_1\omega^{2(p-1)}+\cdots+\alpha_p=0\ (\alpha_i<F)$$

*such that $\phi(\omega)\equiv\psi(\omega^2)$ where $\psi(\rho)=0$ is cyclic. Then $\phi(\omega)=0$ is solvable.*

For since $\psi(\rho)=0$ is a cyclic resolvent equation of $\phi(\omega)=0$, then the adjunction of a root of the former reduces the group $G$ of the latter to an invariant subgroup $G_1$ and the quotient group $H_1=G/G_1$ is cyclic of order $p$. Now $G_1$ is the group of

$$\phi(\omega)\equiv\pi(\omega^2-\rho_i)=0$$

for the enlarged field containing the $\rho_i$ and hence is of order $2^k$. Hence the group of $\phi(\omega)=0$ is of order $2^kp$ and has a series of composition $p, 2, \cdots, 2$. Thus the theorem.

3. **General division algebras.** Let $D$ be a division algebra over any algebraic field $F$, and let it contain a number $i$ whose reduced equation, $\phi(x)=0$ of degree $r$, has at least two roots in $F(i)$. Then this equation has a root in $F(i)$ distinct from $i$, say $\theta(i)$; and hence its distinct symbolic powers $\theta, \theta^2, \cdots, \theta^n=i$ are, also, roots of $\phi(x)=0$. By previous work,* there is a number $J\neq 0$ in $D$ and a polynomial $\theta(i)$ with coefficients in $F$ such that $Ji=\theta(i)J$ if and only if $\theta(i)$ is a root of $\phi(x)=0$. Moreover, corresponding to each root of $\phi(x)=0$ which is in $F(i)$ there corresponds a number $J\neq 0$ in $D$ which is essentially unique in the sense that all such numbers $J$ corresponding to the same root are products of a particular such $J$ by some (any) number in $F(i)$.†

Accordingly, let $J_1, J_2, \cdots, J_l=1$ be a set of non-zero numbers of $D$ such that $J_k i=\theta_k(i)J_k (k=1, \cdots, l)$ where each $\theta_k$ is a polynomial in $i$ and (by the foregoing remarks) is a root of $\phi(x)=0$. Also let the set of $\theta_k (k=1, \cdots, l)$ be such that they form a closed cycle,‡ namely, such that if $\theta_a$ and $\theta_b$ be any two $\theta$'s of the set (coincident or distinct), then $\theta_a\theta_b$ is also in the set. Then we have

LEMMA 1. *Let $A$ be any algebra over $F$ of Dickson's type $A$ (defined as above) and let $i$ be any number $\neq 0$ of $A$ whose reduced equation, of degree $r$, has at least two roots, $i$ and $\theta(i)$, in $F(i)$. Then there is a number $j\neq 0$ of $A$ such that $ji=\theta(i)j$. If $J_1, J_2, \cdots, J_l$ be any set of numbers of $A$ such that $J_k i=\theta_k(i)J_k$ is a polynomial in $i$ where the $\theta$'s form a closed cycle (defined above), then the totality of numbers of the form $B=\sum_k A_k(i)J_k$, where the $A_k$ range over all polynomials in $i$ with coefficients in $F$, form an algebra $B$ of order $rl$.*

For the set of numbers $B$ is closed under addition and subtraction. Also, $J_kJ_t i=J_k\theta_t(i)J_t=\theta_t(\theta_k)J_kJ_t=\theta_p(i)J_kJ_t$ where $\theta_p(i)$ is in the set of $\theta$'s since they form a closed cycle. Hence $J_kJ_t=X(i)J_t$ where $X(i)$ is a polynomial in $i$§ and the set $B$ is an algebra over $F$. It is already known that the set is of order $rl$. Note that $B$ contains $i$ itself since if we let $\theta_1=i$ then $j_1=i$ and thus $j_1 i=i$.

Dickson proved this lemma‖ for the case when the $\theta$'s are symbolic powers of one of them.

By this lemma, there is in the algebra $D$ a subalgebra $D_1$ corresponding

---

* Dickson, these Transactions, vol. 15 (1914), p. 35.

† Hazlett, these Transactions, vol. 18 (1917), p. 171.

‡ This is equivalent to saying that the corresponding substitutions ($H$) of Dickson's paper in the Transactions for April, 1926, form a group.

§ Hazlett, these Transactions, vol. 18 (1917), p. 171.

‖ These Transactions, vol. 15 (1914), p. 36.

to each root of the reduced equation $\phi(x)=0$ of the number $i$ which is in $F(i)$. Since $D$ is an associative division algebra, so also is $D_1$. Moreover, since the algebra $D_1$ is an algebra of the type $A$ discussed by Dickson, his results apply here. Thus a number $x$ of $D_1$ is commutative with all numbers of $D_1$ if $x$ is in the field $K$ generated by the elementary symmetric functions of $\theta_1, \theta_2, \cdots, \theta_l$. Also, $D_1$, an algebra of order $rl$ over $F$, can be regarded as a quadrate algebra $Q_1$ of order $r^2$ and rank $r$ over $K$. Hence $D$ over $F$ contains a quadrate subalgebra $Q_1$ of order $r^2$ defined over $K$, where $K$ is an overfield of $F$ which lies in $D$.

Since $D$ is a quadrate division algebra, there is some overfield $F_1$ of $F$ such that $D$ is equivalent in $F_1$ to a simple matric algebra, $M$. Also, $Q_1$, considered as an algebra over $K$, where $K$ is an algebra over $F_1$, is equivalent to a simple matric algebra $M_1$ which is a subalgebra of $M$. Hence, by one of Wedderburn's theorems, $M$ is the direct product of $M_1$ and a simple matric algebra $M_2$ of order $p^2$, where $(pl)^2$ is the order of $D$. By the foregoing, $M_1$, considered as an algebra over $K$, where $K$ is an algebra over $F_1$, is equivalent in $F_1$ to $Q_1$ over $K$, where $K$ is now an algebra over $F$; but, since $M_2$ contains all numbers of $M$ commutative with every number of $M_1$, this means that the numbers of $K$ come from $M_2$ and $K$ is equivalent in $F_1$ to a subalgebra of $M_2$. Moreover, $M_2$ does not have a basis which, when expressed in terms of the numbers of $D_1$, is rational in $F$. For, if it did, then (by one of Wedderburn's theorems) $M_2$ would be equivalent in $F_1$ to a division algebra $D_2$ over $F$ which is a subalgebra of $D$ and thus* $D$ would be the direct product of $D_2$ and another quadrate division algebra, $D_3$, which contains all numbers of $D$ commutative with every number of $D_2$. Thus $D_3$ would be equivalent to $M_1$ and hence would be equivalent to $D_1$ in $F_1$; and, since $D_1$ and $D_3$ are both subalgebras of $D$, then they would be equivalent in $F$. But this last is impossible since $M_1$ does not have a basis which, when expressed in terms of numbers of $D$, is rational in $F$. Hence $M_2$ does not have a basis rational in $F$.†

Thus we have

REMARK 1. *Let $D$ be a division algebra over the field $F$ and let $D$ contain $Q_1$ when $Q_1$ is a quadrate division algebra defined over $F(x)$ where $x$ is in $D$ but not in $F$. Then $D$ can be represented as a division algebra defined over $Q_1$.*

Applying this to our algebra $D$, we see that if $D$ contain any Dickson algebra $D_1$ as a subalgebra, then $D$ can be represented as a division algebra defined over a quadrate Dickson algebra $Q_1$, where $Q_1$ is defined over $F(\alpha)$, an over-field of $F$. From the foregoing paragraphs, there is such a quadrate algebra $Q_1$ determined by $i$ and any set of polynomials $\theta_1(i), \cdots, \theta_l(i)$ which

* *On division algebras*, these Transactions, vol. 22 (1921), p. 132.

† I here want to thank Professor Wedderburn for pointing out a mistake in this part.

are roots of $\phi(x)=0$ provided the $\theta$'s form a closed cycle. Since the symbolic powers of any such $\theta$ form a closed cycle, we may take these symbolic powers of $\theta$ which are distinct; then they (including $\theta^n=i$) satisfy an equation $g(x)=0$ of degree $n$ with coefficients in $K$ (=the field of the symmetric functions of $\theta, \theta^2, \cdots, \theta^n=i$) which is, moreover, irreducible in $K$.* In this case, $Q_1$ is a quadrate Dickson algebra of order $n^2$ over $K$. If $n$ is composite, let $n=pq$ where $p$ is a prime. Then $\theta^q$ is a polynomial in $i$ with coefficients in $F$ such that its first $p$ symbolic powers are distinct, form a closed cycle, to which the above results apply, and thus are the roots of an equation of degree $p$ with coefficients in $K^1$(= field of the symmetric functions of the symbolic powers of $\theta^q$) which is irreducible in $K$. The quadrate algebra, $Q$, defined by this closed cycle is accordingly a Dickson algebra of order $p^2$ over $K$. Applying Remark 1, we have

REMARK 2. *If the division algebra $D$ has a Dickson proper subalgebra $D_1$, then $D_1$ is a quadrate Dickson algebra $Q_1$ over $F(\alpha)$, where $F(\alpha)$ is an overfield of $F$, and $D$ can be represented as an algebra $E_1$, defined over $D_1$. In its turn, $D_1$ can be regarded as an algebra $E_2$ over an algebra $D_2$, etc., where each of the component algebras $D_i$ is a division algebra (defined over a field $F_i$ contained in $E_i$) of one of the three following types:*

(1) *algebraic fields,*

(2) *normal Dickson algebras,*

(3) *algebras which contain no Dickson subalgebra.*

4. **Solvable algebras.** We now turn our attention to the algebras of Dickson's two most recent papers on division algebras. The above argument applies here.

In the first of these papers,† he is concerned with any associative algebra $\Gamma$ with modulus 1 over any field $F$ with the following properties.

(I) $\Gamma$ is of order $n^2$.

(II) $\Gamma$ contains an element $i$ which satisfies an equation $f(x)=0$ of degree $n$ irreducible in $F$.

(III) The only elements of $\Gamma$ which are commutative with $i$ are polynomials in $i$ with coefficients in $F$.

(IV) The roots of $f(x)=0$ are rational functions, $\theta_r(i)$, of $i$ with coefficients in $F$.

(V) The group, $G$, of $f(x)=0$ is solvable.

(VI) $\Gamma$ contains elements $j_r \neq 0 (r=1, \cdots, n)$ satisfying $j_r i=\theta_r(i) j_r$.

(VII) The product of any $j_r$ and $j_s$ of VI is not zero.

---

* These Transactions, vol. 15 (1914), p. 37.

† *New division algebras*, these Transactions, vol. 28 (1926), pp. 207–234.

Such an algebra he called of Type $E$. For convenience, we might call any algebra of this type a *solvable algebra*. Note that any solvable algebra whose order is the square of a prime is necessarily a Dickson algebra, and that all Cecioni and Dickson algebras are special cases of solvable algebras.

In the following discussion, we shall consider $f(x)=0$ as an ordinary algebraic equation and shall restrict attention to its roots $\theta$ in an overfield of $F(i)$. Let $G$ be the Galois group of order $n$ of the equation $f(x)=0$. If the elements of $G$ be denoted by $\Theta_r$, the notation may be so chosen that $\Theta_r$ is the only substitution that replaces $i$ by $\theta_r(i)$, and also $\Theta_r\Theta_s=\Theta_t$ if and only if $\theta_r\theta_s=\theta_t$. Moreover, since $G$ is solvable, it has an invariant subgroup $G_1$ of prime index $p$ and order $q$, such that every substitution of $G$ can be expressed in the form $\Phi\Theta_k(1\leqq k\leqq q;\ 0\leqq r\leqq p)$, where $\Theta_k$ ranges over the substitutions of $G$, and $\Phi$ is a substitution of $G$ not in $G_1$ and such that the $p$th power of $\Phi$ is the lowest power of $\Phi$ in $G_1$.*

If $\Theta_r$ and $\Theta_s$ be any two substitutions of $G_1$, then $\Theta_t$ is in $G$, and we have $j_k i=\theta_k(i)j_k$ $(1\leqq k\leqq q)$ and $j_r j_s=c_{rs}j_t(c_{rs}<F(i))$. Thus the totality of linear functions of $j_1=1, j_2, \cdots, j_q$ with coefficients in $F(i)$ is a subalgebra, $A_1$, of $\Gamma$ of order $nq=pq^2$ over $F$. Moreover, each number of $A_1$ is commutative with every number of the field $F_1$ of the symmetric functions of $\theta_1=i, \theta_2, \cdots, \theta_q$. Since $F(i)$ is an algebraic field of order $n=pq$ with respect to $F$ and of order $q$ with respect to $F_1$, then $F_1$ is of order $p$ with respect to $F$ and is defined by an irreducible equation $g_1(x)=0$ of degree $p$, with coefficients in $F$. Thus $A_1$ is a subalgebra $S_1$ of order $q^2$ and rank $q$ over $F_1$.

From this it follows that we can arrange the $n$ roots of $f(x)=0$ in $p$ rows of $q$ each and arrange similarly the corresponding solutions $j\neq 0$ of the equations $ji=\theta j$ as in the accompanying array:

$$\text{I}\qquad \left\{\begin{array}{llll}
\phi_1(\theta_1)=i, & \phi_1(\theta_2)=\theta_2, & \cdots, & \phi_1(\theta_q)=\theta_q \\
j_1J_1=j_1=J_1=1, & j_2J_1=j_2, & \cdots, & j_qJ_1=j_q \\
\phi_2(\theta_1)=\phi_2, & \phi_2(\theta_2), & \cdots, & \phi_2(\theta_q) \\
j_1J_2=J_2, & j_2J_2, & \cdots, & j_qJ_2 \\
\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots & \cdots\cdots & \cdots\cdots \\
\phi_p(\theta_1)=\phi_p, & \phi_p(\theta_2), & \cdots, & \phi_p(\theta_q) \\
j_1J_p=J_p, & j_2J_p, & \cdots, & j_qJ_p
\end{array}\right.$$

where we have placed each solution $j\neq 0$ directly below the corresponding $\theta$. The $q$ $\theta$'s in the first line are the roots of $f_1(x)=0$ of degree $q$, the $q$ $\theta$'s in the second line are the roots of an equation $f_1'(x)=0$ obtained by replacing $i$ in

* For details, see the paper by Dickson cited in the previous footnote.

the coefficients of $f_1(x)=0$ by $\phi_2(i)$; etc. Since $\Phi^p<G$, then $\phi^p$ is a root in the first row, but is not $i$ unless $\Phi=1$.

Now $G_1$ is the group of $f_1(x)=0$ for the field $F_1$. For, if we adjoin to $F$ a root of $g_1(x)=0$, this is equivalent to enlarging $F$ to $F_1$ and the group is reduced to an invariant subgroup of $G$. Since any symmetric function of the roots of $f_1(x)=0$ is unaltered by every transformation of $G_1$, then the group contains $G_1$ and thus is precisely $G_1$.

From this, it follows that the group of $g_1(x)=0$ for the field $F$ is the quotient group $G/G_1=H_1$ and $g_1(x)=0$ is a cyclic equation. Since $g_1(x)=0$ is completely solvable in $F(i)$, we may denote any of its roots by $(i)$. Since $g_1(\psi(x))=0$ has the root $i$ in common with the irreducible equation $f(x)=0$, then it has all the roots of the latter and thus $g_1(x)=0$ has $\psi(\theta)$ as a root where $\theta$ is any root of $f(x)=0$. If $h(x)=\pi(x-\psi(\theta_k))$ $(1\leqq k\leqq n)$, then every coefficient of $h(x)$ is a symmetric function of the roots of $f(x)=0$ and thus is in $F$. Moreover, every root of $h(x)=0$ is a root of the irreducible equation $g_1(x)=0$. Thus every root of $g_1(x)=0$ is of the form $\psi(\theta)$, and $h(x)$ is the $q$th power of $g_1(x)$.

From this relation of the roots of $g_1(x)=0$ to the roots of $f(x)=0$, we readily obtain a set of numbers $j$ of the algebra $\Gamma$ satisfying the equation $j\psi(i)=\chi(i)j$. For if $\theta$ be any root of $f(x)=0$ and $j\neq 0$ a solution of $ji=\theta j$, then $j\,\psi\,(i)=\psi(\theta)j$. Moreover, if $\chi(i)$ be any number of $F$ such that there exists a number $J\neq 0$ satisfying $J\psi(i)=\chi(i)J$, then $0=Jg_1(\psi)=g_1(\chi)J$ and hence $\chi$ is a root of $g_1(x)=0$. Accordingly, all such polymonials $\chi$ are of the form $\psi(\theta)$, where $\theta$ is a root of $f(x)=0$; and we find a solution $J\neq 0$ in $\Gamma$ of each possible equation $J\psi=\chi J$ by taking as $\chi$ in turn every polynomial of the form $\psi(\theta)$ and then a corresponding $J$ is the $j$ corresponding to $\theta$ in the equation to $ji=\theta j$.

It is to be noted that, although a number $j$ in $\Gamma$ such that $ji=\theta j$ is essentially unique* in the sense that any other $j$ satisfying the equation is of the form $\beta(i)j$, yet a number $j$ in $\Gamma$ such that $j\psi(i)=\psi(\theta)j$ is not essentially unique. For, since $f(x)=0$ has $n$ roots, there are $n=pq$ essentially distinct numbers $j$ that satisfy in turn the $p$ possible equations $j\psi=\chi j=\psi(\theta)j$. Since $h(x)$ is the $q$th power of $g_1(x)$, the roots $\psi(\theta)$ of $h(x)=0$ are equal in sets of $q$ each but the $q$ corresponding $j$'s are not essentially equal. In fact, $\psi(i)$ is a polynomial in $i$ which is a symmetric function of the roots $\theta$ in the first row of the array I and the distinct roots of $g_1(x)=0$ are respectively the $\psi$-functions of each of the roots in the first column, and thus each $\psi(\phi_k)$ is a symmetric function of the roots in the $k$th row. Thus we may choose as the numbers $j$ corresponding to the roots of $g_1(x)=0$ any $q$ numbers such that one

* Since, under the assumptions, $\Gamma$ has the basis $i^k j_r (k, r=1, \cdots, n)$.

is chosen from each row. That is, we may choose the $j$'s of the first column; and sometimes we may so choose them that the corresponding $\theta$'s form a closed cycle. In the latter case, these $j$'s are closed under multiplication except, possibly, for multipliers in $F(i)$. In any case, denote the $j$'s by $J_1=1, J_2, \cdots, J_q$. The totality of numbers

$$\sum_{l,t} a_{lt}[\psi]^l J_t \qquad (a<F;\ l, t=1, \cdots, p)$$

will be denoted by $\Gamma_1$.

Then the set of numbers $j$ corresponding to the roots of $f(x)=0$ is obtained by forming all possible products $J_k j_l (k=1, \cdots, p; l=1, \cdots, q)$ or, if we wish, by forming all possible products $j_l J_k$. Every number of $\Gamma$ is of the form

$$\sum_{s,t} \beta_{st}(i) j_s J_t \qquad (s=1, \cdots, q; t=1, \cdots, p),$$

where each $\beta<F(i)$ and therefore is of the form

$$\sum_{k,l} \alpha_{kl} i^k [\psi]^l \qquad (k=1, \cdots, q;\ l=1, \cdots, p).$$

Hence every number of $\Gamma$ is of the form

$$\sum \beta_{klst}(i^k j_s)([\psi]^l J_t),$$

where each $\beta<F$. Let $A_1$ denote the totality of numbers of the form

$$\sum_{k,s} b_{ks} i^k j^s \qquad (k, s, =1, \cdots q).$$

Note that, although the foregoing was proved under the assumption that $p$ is a prime, yet the same reasoning holds if $G_1$ be any invariant subgroup of index $p$ under $G$, a solvable group, since a resolvent equation of a solvable equation is always solvable.

Since (1) every number linearly dependent with respect to $F$ on the products $i^k j_s (k, s=1, \cdots, q)$ is a number of $A_1$ and conversely every number of $A_1$ can be expressed in one and only one way as such a linear combination; (2) every number linearly dependent with respect to $F$ on the products $[\psi]^l J_t$ $(l, t=1, \cdots, p)$ is a number of $\Gamma_1$ and conversely, we shall say that $\Gamma$ over $F$ has been represented as the algebra $\Gamma_1$ over the algebra $A_1$, where $A_1$ is defined over $F$.

If the $J$'s can be so chosen that the set is closed under multiplication except for multipliers in $F_1$, then $\Gamma_1$ is an algebra of order $p^2$ over $F$. Moreover, $\Gamma_1$ is a normal algebra, for if $x=\sum x_{kl}[\psi]^k J_l (x_{kl}<F)$ be any number of $\Gamma_1$ that is commutative with every number of $\Gamma_1$, then $x\psi=\psi x$ is equivalent to

$$\sum_{k,l} x_{kl}[\psi]^k \psi_l J_l = \sum_{k,l} x_{kl}[\psi]^{k+1} J_l,$$

where $J_l\psi = \psi_l J_l$, and this is equivalent to

$$\sum_k x_{kl}[\psi]^k(\psi_l - \psi) = 0.$$

For each value of $l$ there are two possibilities: $\psi = \psi_l$ or $\sum_k x_{kl}[\psi]^k = 0$. But $\psi \neq \psi_l$ unless $l = 1$, and hence $\sum x_{kl}[\psi]^k = 0$ $(l \neq 1)$, so that $x < F(\psi) = F_1$. Furthermore, $x(\psi)J_r = J_r x(\psi)$ if and only if

$$\sum_k x_k([\psi]^k - [\psi_r]^k)J_r = 0, \quad x_k([\psi]^k - [\psi_r]^k) = 0,$$

which is equivalent to saying that $x_k = 0$ unless $[\psi]^k = [\psi_r]^k$ for every $r$. But $\psi \neq \psi_r (r \neq 1)$ and thus $x_k = 0 (k \neq 0)$, so that $x < F$.

Since $\Gamma_1$ is quadrate and is defined over the same field $F$ as $\Gamma$, then, by by one of Wedderburn's theorems, $\Gamma$ is the direct product of $\Gamma_1$ and another quadrate division algebra, $\Gamma_2$, over $F$. To determine $\Gamma_2$, we must determine the totality of numbers of $\Gamma$ that are commutative with every number of $\Gamma_1$. Now every number of $\Gamma$ is of the form $x = \sum_{st} x_{st}(i) j_s J_t$. Accordingly, $x\psi = \psi x$ is equivalent to

$$\sum_{st} x_{st}(i)\psi(\phi_t(\theta_s)) j_s J_t = \sum_{st} x_{st}(i)\psi j_s J_t,$$

$$x_{st}(i)[\psi(\phi_t(\theta_s)) - \psi] = 0.$$

Thus $x_{st} = 0$ unless $\psi(\phi_t(\theta_s)) = \psi$, and this last is equivalent to $\psi_t \equiv \psi(\phi_t) = \psi(\theta_t) = \psi$, where $\theta_s(\theta_t) = i$. Thus $x_{st} = 0$ unless $t = 1$, so that $x = \sum_s x_s(i) j_s$.* But the totality of such numbers $x$ contains all numbers of $F(i)$. Now if $x < F(i)$, then $xJ_r = J_r x$ if and only if $x(i)J_r = x(\phi_r)J_r$, and thus such an $x$ is commutative with every $J_r$ if and only if $x < F_2$, the field of symmetric functions of the $\phi_r$. Just as we proved that $F_1$ is of order $p$, so we prove that $F_2$ is of order $q$.

Since $\Gamma_2$ is a quadrate division algebra of order $q^2$ containing the number $\chi$ defining $F_2$ which satisfies an irreducible solvable equation of degree $q$, then $\Gamma_2$ is a solvable algebra and contains a set of numbers $K_l (l = 1, \cdots, q)$ such that $K_l\chi = \rho_l(\chi)K_l$ where $\rho_l$ is a polynomial in $\chi$ with coefficients in $F$, and such that every number of $\Gamma_2$ is of the form $\sum y_l K_l (l = 1, \cdots, q)$ where $y_l < F_2$. Since $\rho_l(\chi)$ is a conjugate of $\chi$ with respect to $F$, then each $\rho$ is of the form $\chi(\theta)$, where $\theta$ is a conjugate of $i$ with respect to $F$. But $\chi$ is a symmetric function of the roots of $f(x) = 0$ in the first column of I, and hence (by the argument used above for the conjugates of $\psi$) the conjugates of $\chi$

* Note that, even if $\Gamma_1$ be not an algebra, the totality of such numbers $x$ is precisely the totality of numbers of $\Gamma$ commutative with the numbers of $F_1$.

are the same symmetric functions of the roots in the second column, in the third column, and so on. If $\chi \equiv \chi(i)$, then the conjugates of $\chi$ are $\chi(\theta_l)$ $(l=1, \cdots, q)$. Hence $K_l\chi(i)=\chi(\theta_l)K_l=\chi_lK_l$.

But $K_l=\sum_s x_s(i)j_s$, and this last is equivalent to

$$x_s\chi(\theta_s) = x_s\chi(\theta_l) \qquad (s = 1, \cdots, q)$$

which is equivalent to $\chi_s=\chi_l$ if $x_s\neq 0$. Since the $q$ conjugates of $\chi$ are distinct, then $x_s=0$ unless $s=l$, and thus

$$K_l = x_l(i)j_l \qquad (l = 1, \cdots, q).$$

Since $K_l\neq 0$ and $K_li=\theta_lK_l$, we may take $K_l$ in place of the old $j_l$, and we shall assume henceforth that we have so done. With this choice of $j_1, \cdots, j_q$, $\Gamma_2$ is the totality of numbers of the form $\sum_s x_sj_s$, where $x_s<F_2$. Note that the new set of $j$'s is closed under multiplication except for the multipliers in $F_2$.

Thus we have the

THEOREM. *Let $\Gamma$ be a solvable algebra of order $n^2$ over a field $F$. Let $i$ be any number of $\Gamma$ satisfying a solvable irreducible equation $f(x)=0$ of degree $n$ with roots in $F(i)$ as given by the square array* I *and let the essentially unique number $j\neq 0$ of $\Gamma$ satisfying the equation $ji=\theta j$ corresponding to each such root $\theta$ be as given directly below $\theta$ in* I. *Let $F_1$ be the field of symmetric functions of $\theta_1, \cdots, \theta_q$ of the first row, and let it be defined by $\psi$ where $\psi$ satisfies an irreducible equation in $F$ of degree $p$. Assume that the set of numbers $J_1, \cdots, J_p$ corresponding to the roots of the first column is closed under multiplication except for multipliers in $F_1$. Then the totality of numbers $\sum_{lt}a_{lt}[\psi]^tJ_l(a<F)$ form a solvable division algebra $\Gamma_1$ of order $p^2$. Also, the set of roots $\phi$ of the first column is closed under iteration and their symmetric functions determine a field $F_2=F(\chi)$ defined by an irreducible equation of degree $q$. If the set of numbers $j_1, \cdots, j_q$ is not closed under multiplication except for multipliers in $F_2$, then each $j_s$ may be multiplied by a suitable number $x_s(i)\neq 0$ such that the set $x_1j_1, \cdots, x_qj_q$ is closed under multiplication except for multipliers in $F_2$. If we take these as a new set of $j_1, \cdots, j_q$ then the totality of numbers of the form $\sum_{ks}b_{ks}[\chi]^kj_s$ form a solvable division algebra $\Gamma_2$ of order $q^2$. The algebra $\Gamma$ is the direct product of $\Gamma_1$ and $\Gamma_2$.*

It should be noted that the assumption that the set of numbers $J_1, \cdots, J_p$ corresponding to the roots of the first column is closed under multiplication except for multipliers in $F_1$, is not as strong as it may at first seem. For if the set of numbers $J_1, \cdots, J_p$ of the first column is not closed under multiplication, there may be $p$ non-zero numbers in $F(i)$, $y_1, \cdots, y_p$,

such that the set $y_1J_1, \cdots, y_pJ_p$ is closed under multiplication except for multipliers in $F_1$. Moreover, even if this be not possible it may be possible for another set of $J_1=1, \cdots, J_p$; for $J_2$ is any non-zero solution $j$ of the equation $ji=\theta j$ where $\theta$ is any root of $f(x)=0$ not in the first row; $J_3$ is any non-zero solution when $\theta$ is any root not in the first two rows; etc.

Moreover, this assumption as to the closure of the set of $J$'s is not incompatible with the assumption that $\Gamma$ is an associative solvable algebra, as will be seen if we turn to Dickson's article.* In §1, he exhibits a class of solvable division algebras of order $p^2q^2$ over $F$ with basal numbers $i^aj^bk^c$ $(a<pq,\ b<q,\ c<p)$, where $i$ satisfies an irreducible equation of degree $pq$ with roots $\theta^k[\psi^r(i)]=\psi^r[\theta^k(i)]$ $(k=0,\ 1, \cdots, q-1;\ r=0,\ 1, \cdots, p-1)$ where $\theta^q=i$, $\psi^p=i$. Multiplication is determined by the associative law in conjunction with $j^q=g$, $k^p=r$, $kj=\alpha jk$, $ji=\theta j$, $ki=\psi k$ where $q$, $r$, $\alpha<F(i)$. The conditions for associativity reduce to

$$g = g(\theta),\quad \alpha\alpha(\theta)\alpha(\theta^2)\cdots\alpha(\theta^{q-1})g = g(\psi),$$

$$r = r(\psi),\quad \alpha\alpha(\psi)\alpha(\psi^2)\cdots\alpha(\psi^{p-1})r(\theta) = r.$$

Here the powers of $j$ play the rôles of $j_1, \cdots, j_q$ of the present paper and the powers of $k$ play the rôles of $J_1, \cdots, J_p$. To see that our assumption as to the set of the $J$'s is not inconsistent with the above condition for associativity, take $g$ and $r$ as numbers in $F$ and then both the sets of $j$ and those of $J$ are closed under multiplication. From the work above, it follows that $kj=jk$ and therefore $\alpha=1$ and all conditions for associativity are satisfied.

Brauer† has proved that any quadrate division algebra of order $p^2q^2$ where $p$ and $q$ are relatively prime is the direct product of a division algebra of order $p^2$ and a division algebra of order $q^2$. In particular, any solvable algebra $\Gamma$ of order $p^2q^2$ is a direct product of a division algebra $\Gamma_1$ of order $p^2$ and a division algebra $\Gamma_2$ of order $q^2$. If $\Gamma_1$ be solvable, then it contains a number $i$, which satisfies an irreducible solvable equation in $F$ of degree $p$; and similarly with $\Gamma_2$. Applying Theorem 21 of A. A. Albert's first paper‡ or by Theorem 2 of our footnotes re his theorem (see our §2), it follows that the $i$ specified above may be taken as any number defining $F(i_1, i_2)$. As the resolvent equation $g_1(x)=0$, we shall take any defining equation of $F(i_1)$. Using the notation of the proof of the theorem, we first have the numbers $J_1, \cdots, J_p$ given us in $\Gamma_1$, and thus we have $J_li=\alpha_l(i_1)J_l$. But $i_l$ is a polynomial in $i$, say $\psi(i)$; and hence, by the proof of the theorem, $\alpha_l(i)=\psi(\phi_l)$

* These Transactions, vol. 28, p. 207.

† Loc. cit., p. 104.

‡ Loc. cit., p. 329.

where $\phi_l$ is some root of $f(x)=0$, the irreducible equation defining $i$. We may take the present $\phi_2, \phi_3, \phi_4, \cdots$ as the $\phi_2, \phi_3, \phi_4, \cdots$ of I. For the $\alpha_l$ combine under iteration in a manner that might be described as simply isomorphic with the manner in which the corresponding elements of the group $H_1$ of $g_1(x)=0$ combine under multiplication, and the corresponding $J_l$ combine in the same manner, so to speak, except that the order of multiplication is reversed.* Since $H_1$ is the quotient group, there is an $\alpha_l$ corresponding to each row of I, this correspondence applying to multiplication; also, there is a $J_l$ corresponding to each row of I and this $J_l$ may be taken as one of the $j$'s in that row of I. It is only a matter of rearranging the elements of one row to put this element first in the row. Hence the present $J_1, \cdots, J_p$ can be taken as the $J_1, \cdots, J_p$ of I and then the present $\phi_1, \cdots, \phi_p$ become the $\phi$'s of I. Since the set of $J$'s is closed under multiplication except for multipliers in $F_1=F(i)=F(\psi)$, we see that the assumptions of our theorem are satisfied whenever $\Gamma$ is the direct product of two solvable division algebras.

* For details, see array I.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

# ON THE STRUCTURE OF CONNECTED AND CONNECTED IM KLEINEN POINT SETS*

BY

G. T. WHYBURN

## 1. INTRODUCTION; NOTATION; SOME KNOWN RESULTS

In this paper it is shown that all connected and connected im kleinen point sets lying in a separable metric space admit of decomposition into elements, called nodular elements, which are highly analogous to the cyclic elements of a continuous curve† and that the structure of these sets relative to their nodular elements is very similar to the structure of continuous curves relative to their cyclic elements. Indeed, it is true that for the special case of a continuous curve $M$, the nodular elements of $M$ are identically the same as the cyclic elements of $M$.

The usual terminology and notation of point set theory will be employed. To facilitate the reading of the paper, however, some of the less familiar terms and notations used are here defined and explained. The point $P$ of a connected point set $M$ is a non-cut point or a cut point of $M$ according as $M-P$ is or is not connected. The subset $X$ of a connected set $M$ separates two subsets $A$ and $B$ of $M$ in $M$ provided that $M-X$ is the sum of two mutually separated sets $M_a(X)$ and $M_b(X)$ containing $A$ and $B$ respectively, and this fact is indicated by means of the equation $M-X=M_a(X)+M_b(X)$. The subset $K$ of a point set $M$ is said to be closed relative to $M$, or closed in $M$, provided that no point of $M-K$ is a limit point of $K$. The symbol $\supset$ means "contains," and $\subset$ means "is contained in"; $\delta(H)$ denotes the diameter of the point set $H$, and $\rho(X, Y)$ denotes the lower bound of the aggregate of numbers $[\rho(x, y)]$, where $x$ and $y$ are points of $X$ and $Y$ respectively and $\rho(x, y)$ is the distance from $x$ to $y$.

Since this paper is concerned very largely with connected and connected im kleinen sets, it is to be understood, unless definitely stated to the contrary, that when we speak of a "set $M$" it is presupposed that $M$ is a connected and connected im kleinen point set. The following results, some of which will be used later, either are already known or follow, as indicated below, quite readily from known properties of connected im kleinen point sets.

* Presented to the Society, December 27, 1928; received by the editors in October, 1929.

† See my paper *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), pp. 167–194.

(1.1) *If $A$ and $B$ are two points of a set $M$ and $K$ is the set of all those points of $M$ which separate $A$ and $B$ in $M$, then $K+A+B$ is a closed and compact point set.*

(1.2) *Using the same notation as in* (1.1), *with $M$ any connected point set, the points of $K$ are linearly ordered from $A$ to $B$ in $M$; and if $H$ is any closed and compact subset of $K$, there exists a first point of $H$ in the order from $A$ to $B$.*

(1.3) *The set $G$ of all the cut points of a set $M$ is an $F_\sigma$ set, i.e., the sum of a countable number of closed sets.*

(1.4) *If the two points $A$ and $B$ of a set $M$ are not separated in $M$ by a subset $X$ of $M$ which is closed in $M$ and contains neither $A$ nor $B$, then $A$ and $B$ lie together in a connected subset of $M-X$.*

Result (1.1) has been proved by the author for a less general space elsewhere.* The proof there given for the closure of the set $K+A+B$ holds equally well in the space considered in the present paper. That $K+A+B$ is compact is proved in general as follows: Suppose, on the contrary, that $K+A+B$ contains an infinite subset $D$ having no limit point. Now for each point $X$ of $K$, $M-X=M_a(X)+M_b(X)$, where the sets $M_a(X)$ and $M_b(X)$ are mutually separated and contain $A$ and $B$ respectively. As shown in my paper just cited, the set $D$ contains a sequence $X_1, X_2, X_3, \cdots$ of distinct points of $K$ such that either for each $i$, $X_i \subset M_b(X_{i-1})$, or for each $i$, $X_i \subset M_a(X_{i-1})$. The two cases are alike, so we shall suppose the former. Now the set of points $H=\sum_1^\infty M_a(X_i)$ is closed in $M$, because the set $\sum_1^\infty X_i$ is closed and is a subset of $H$, and $M$, by hypothesis, is connected im kleinen. But $H$ is also open in $M$, since the sets $M_a(X_i)$ are open in $M$. This is impossible, because $H \neq M$ and $M$ is connected.

The first part of result (1.2) has also been proved by the author elsewhere.† The second part of this result follows readily from the first part of (1.2) and the fact that $H$ is closed, compact, and separable. Result (1.3) follows from (1.1) and the fact that $M$ is separable by an argument similar to, but simpler than, that used by Zarankiewicz‡ to prove the same theorem for the special case where $M$ is a continuous curve. I note here the fact that

* See *Concerning connected and regular point sets*, Bulletin of the American Mathematical Society, vol. 33 (1927), p. 685. For a very simple proof of the same theorem, the closure part of which also is valid in the space considered in this paper, see R. L. Wilder, *On connected and regular point sets*, ibid., vol. 34 (1928), pp. 649–655.

† See *Concerning collections of cuttings of connected point sets*, Bulletin of the American Mathematical Society, vol. 35 (1929), pp. 90, 91.

‡ See *Sur les points de division dans les ensembles connexes*, Fundamenta Mathematicae, vol. 9 (1927).

it is also true that the set $K$ of all the *im kleinen cut points* of a set $M$ (i.e., all points $P$ of $M$ such that $P$ is a cut point of some connected open subset of $M$) is an $F_\sigma$ set. Result (1.4) is due to R. L. Wilder.*

## 2. Preliminary lemmas

(2.1) *If $K$ is any closed (in $M$) subset of a connected and connected im kleinen point set $M$ having the property that, for each component $N$ of $M-K$, $\overline{N}\cdot K$ consists of a single point, then for every connected subset $H$ of $M$, $H\cdot K$ is either vacuous or connected.*

Suppose, on the contrary, that, for some connected subset $H$ of $M$, $H\cdot K=H_1+H_2$, where $H_1$ and $H_2$ are mutually separated point sets. Let $G$ be the collection of all those components of $M-K$ each of which contains at least one point of $H$. For each element $g$ of $G$, it is easily seen that the point $P_g=\bar{g}\cdot K$ is a point of $H\cdot K$. Let $G_1$ be the collection of elements $g_1$ of $G$ such that $P_{g_1}\subset H_1$ and $G_2$ the collection of elements $g_2$ of $G$ such that $P_{g_2}\subset H_2$. Let

$$M_1 = H_1 + H\cdot \sum_{g_{1i} \text{ in } G_1} g_{1i} \text{ and } M_2 = H_2 + H\cdot \sum_{g_{2i} \text{ in } G_2} g_{2i}.$$

Then clearly $M_1$ and $M_2$ are mutually exclusive and $M_1+M_2=H$. They must also be mutually separated. For suppose one of them contains a limit point of the other, say $M_1$ contains a limit point $P$ of $M_2$. Now since $M$ is connected im kleinen and $K$ is closed in $M$, therefore each component of $M-K$ is an open subset of $M$; and thus it follows that $P$ does not belong to $H\cdot\sum g_{1i}$. Therefore $P\subset H_1$. Since $H_1$ and $H_2$ are mutually separated, there exists a neighborhood $R$ of $P$ such that $\overline{R}\cdot H_2=0$. But now since $M$ is connected im kleinen and $P$ is a limit point of $M_2$, there exists a point $Q$ of $M_2$ belonging to a component $g_2$ of $M-K$, which lies together with $P$ in a connected subset $I$ of $M\cdot R$. It readily follows that $R\cdot K$ contains a point $X$ which is a limit point of $g_2$. But $P_{g_2}$ does not belong to $R$, and by hypothesis $P_{g_2}$ is the only point in $\bar{g}_2\cdot K$. Thus the supposition that $H\cdot K$ is not connected leads to a contradiction.

(2.2) *Under the conditions of* (2.1), *$K$ is connected and connected im kleinen.*

By letting $M=H$ in (2.1), we get that $M\cdot K=K$ is connected. If $P$ is any point of $K$ and $\epsilon>0$, then since $M$ is connected im kleinen, a $\delta_\epsilon>0$ exists such that each point $Q$ of $K$ with $\rho(Q, P)<\delta_\epsilon$ lies together with $P$ in a con-

* See *A characterization of continuous curves by a property of their open subsets*, Fundamenta Mathematicae, vol. 11 (1928), pp. 127–131.

nected subset $I$ of $M$ with $\delta(I)<\epsilon$. Then since, by (2.1), $I\cdot K$ is connected and contains $P+Q$, it follows that $K$ is connected im kleinen at each of its points $P$.

For convenience of reference we state here also the following easily established lemma which was stated to me in less general form by R. L. Moore.

(2.3) *If $H$ is any subset of a connected point set $M$, then the set $K$ of all those points of $M$ which are not separated in $M$ from any point of $H$ by any single point of $M$ is closed in $M$.*

## 3. Nodular subsets

**Definitions.** A connected subset $C$ of a connected point set $M$ will be called "nodular" provided that no point of $C$ is a cut point of $C$. A nodular subset $C$ of $M$ is called a maximal nodular subset of $M$ provided that $C$ is saturated with respect to the property of being a nodular subset of $M$.

(3.1) *Every maximal nodular subset $C$ of a connected and connected im kleinen set $M$ is closed (in $M$) and connected.*

By definition, $C$ is connected. Suppose $P$ is any limit point of $C$ which belongs to $M$. Then $P$ must belong to $C$. For since $C$ has no cut point, obviously $C+P$ has no cut point. Then $C+P$ is a nodular subset of $M$ containing $C$ as a proper subset, contrary to the fact that $C$ is a maximal nodular subset of $M$.

(3.2) *A nodular subset $C$ of a connected and connected im kleinen set $M$ is maximal if $C$ is closed (in $M$) and no component of $M-C$ has more than one limit point in $C$.*

Suppose, on the contrary, that $C$ is a proper subset of some nodular subset $C_0$ of $M$. Let $N$ be a component of $M-C$ containing a point $P$ of $C_0-C$. Since $C$ is closed and $M$ is connected im kleinen, $N$ must have at least one limit point $Q$ in $C$; and by hypothesis $Q$ is the only limit point of $N$ in $C$. Then, using the connectivity im kleinen of $M$ it follows that $Q$ belongs to $C_0$ and that $C_0-Q$ is the sum of the two mutually separated sets $N\cdot C_0$ and $C_0[M-(N+Q)]$. Thus $Q$ is a cut point of $C_0$, contrary to the fact that $C_0$ is a nodular subset of $M$.

(3.3) *If a subset $C$ of a connected and connected im kleinen set $M$ contains a non-cut point $P$ of $M$, then $C$ is a maximal nodular subset of $M$ if and only if $C$ is the set of all those points of $M$ which are not separated in $M$ from $P$ by any point of $M$.*

The condition is sufficient. By (2.3), $C$ is closed in $M$. I shall now show

that each component of $M-C$ has not more than one limit point in $C$. Suppose, on the contrary, that some component $N$ of $M-C$ has two limit points $A$ and $B$ in $C$. Since $M$ is connected im kleinen, it easily follows that there exist two connected and connected im kleinen subsets $I_a$ and $I_b$ of $M$ containing $A$ and $B$ respectively and such that $I_a \cdot N \neq 0 \neq I_b \cdot N$ and $I_a \cdot I_b = 0$. There exist components $H_x$ and $H_y$ respectively of $I_a \cdot N$ and $I_b \cdot N$ having limit points $X$ and $Y$ respectively in $C$. Since $M$ is connected im kleinen, it follows that $N$ also is connected im kleinen; and if $H$ denotes the component of $N-\overline{H}_x \cdot N$ which contains $H_y$, then $\overline{H}_x \cdot N$ contains a point $Q$ which is a limit point of $H$. Now since $Q$ is not in $C$, there exists a point $Z$ of $M$ which separates $Q$ from $P$ in $M$; $Z$ does not separate either $X$ or $Y$ from $P$ in $M$, because $X+Y \subset C$. Hence $Z$ must separate $Q$ from both $X$ and $Y$ in $M$. But clearly this is impossible, since $H_x+X+Q$ and $H+Y+Q$ are connected subsets of $M$ having only the point $Q$ in common. Thus the supposition that some component of $M-C$ has more than one limit point in $C$ leads to a contradiction.

Thus $C$ has the property mentioned in (2.1), and hence, by (2.2), $C$ is connected. Now $C$ can have no cut point. For suppose $C$ has a cut point $A$. Then if $A \neq P$, $A$ separates some point $Q$ of $C$ from $P$ in $C$; but since $A$ does not separate $Q$ from $P$ in $M$ (for $Q+P \subset C$), and $M$ is connected im kleinen, by (1.4) there exists a connected subset $H$ of $M-A$ containing both $Q$ and $P$. But by (2.1), $C \cdot H$ is connected; and since it contains both $P$ and $Q$, this contradicts the fact that $A$ separates $P$ and $Q$ in $C$. If $A=P$, then $P$ separates some two points $X$ and $Y$ of $C$ in $C$. But by hypothesis $P$ is not a cut point of $M$, and hence it does not separate $X$ and $Y$ in $M$. Hence, by (1.4), $M-P$ contains a connected subset $H$ containing both $X$ and $Y$. But by (2.1), $H \cdot C$ is connected; and since $H \cdot C \supset X+Y$, $P$ does not separate $X$ and $Y$ in $C$. Thus, in any case, the supposition that $C$ has a cut point leads to a contradiction. Therefore $C$ is a nodular subset of $M$. And since, as shown above, no component of $M-C$ has more than one limit point in $C$, it follows by (3.2) that $C$ is a maximal nodular subset of $M$.

The condition is also necessary. For let $C$ be any maximal nodular subset of $M$ containing a non-cut point $P$ of $M$. Let $C_0$ be the set of all those points of $M$ which are not separated in $M$ from $P$ by any point of $M$. Since $C$ is nodular, obviously $C \subset C_0$. Now it was just shown above that $C_0$ is a nodular subset of $M$. Therefore $C$ is not a proper subset of $C_0$, and hence $C=C_0$.

**Note.** In the first paragraph of the proof of (3.3) no use was made of the fact that $P$ is a non-cut point of $M$. Thus we have the following corollary.

(3.3a) *If $P$ is any point of a connected and connected im kleinen set $M$ and $K$*

*is the set of all those points of $M$ not separated in $M$ from $P$ by any point of $M$, then $K$ has the property mentioned in* (2.1). *Hence $K$ is closed in $M$, connected, and connected im kleinen.*

(3.4) *A nodular subset $C$ of a set $M$ is maximal if and only if $C$ is closed in $M$ and no component of $M-C$ has more than one limit point in $C$.*

The condition is sufficient by (3.2). That it is necessary was proved incidentally in the proof of (3.3).

(3.5) *Every maximal nodular subset $C$ of a set $M$ is closed in $M$, connected, and connected im kleinen; and $H \cdot C$ is either vacuous or connected, for every connected subset $H$ of $M$.*

This follows at once from results (2.1), (2.2), (3.1), and (3.4).

(3.6) *If $C_0$ contains more than one point and is any nodular subset of a set $M$, then there exists a maximal nodular subset $C$ of $M$ which contains $C_0$.*

Since $C_0$ is connected and has no cut point, then* it can contain at most a countable number of cut points of $M$. Therefore it contains a non-cut point $P$ of $M$. And if $C$ denotes the set of all those points of $M$ which are not separated in $M$ by any point of $M$, then by (3.3), $C$ is a maximal nodular subset of $M$. Clearly $C_0 \subset C$, because $C_0$ is nodular.

**Definition.** A point $P$ of a maximal nodular subset $C$ of a connected and connected im kleinen point set $M$ which is a non-cut point of $M$ will be called an *internal* point of $C$.

(3.7) *No two maximal nodular subsets of a connected and connected im kleinen set $M$ can have more than one common point, nor can they have an internal point of either in common without being identical.*

This follows at once from result (3.3).

(3.8) *If $A$ and $B$ are any two points of a connected and connected im kleinen point set $M$ which are not separated in $M$ by any point of $M$, then there exists a maximal nodular subset of $M$ containing both $A$ and $B$.*

Let $C$ be the set of all those points of $M$ which are not separated in $M$ from either $A$ or $B$ by any point of $M$. By (2.3), $C$ is closed in $M$. Clearly $C \supset A+B$. And by the same argument as given in the first paragraph of the proof of (3.3), with very slight modifications, it is shown that $C$ has the pro-

* See R. L. Moore, *Concerning cut-points of continuous curves and of other closed and connected point-sets*, Proceedings of the National Academy of Sciences, vol. 9 (1923), pp. 101–106; also see C. Zarankiewicz, loc. cit.

perty mentioned in (2.1). Hence by (2.2), $C$ is connected. Now $C$ can have no cut point. For suppose $X$ is a cut point of $C$. Then $X$ separates some two points $U$ and $V$ of $C$ in $C$. Now $X$ cannot separate $U$ and $V$ in $M$; for if so, then $M-X=M_u(X)+M_v(X)$, and at least one of the sets $M_u(X)$ and $M_v(X)$ must contain a point of $A+B$; but then either $U$ or $V$ is separated in $M$ from one of the points $A$ and $B$ by the point $X$, contrary to the fact that $U+V \subset C$. Hence, by (1.4), $M-X$ contains a connected set $H$ containing $U+V$. But by (2.1), $C \cdot H$ is connected. Clearly this contradicts the supposition that $X$ separates $U$ and $V$ in $C$. Therefore $C$ has no cut point, and hence is a nodular subset of $M$. That $C$ is a maximal nodular subset of $M$ follows now from (3.4) and the fact that $C$ has the property of (2.1).

## 4. Nodular elements; nodules

**Definition.** The point $P$ of a connected and connected im kleinen point set $M$ will be called an end point of $M$ provided there exists an infinite sequence $X_1, X_2, X_3, \cdots$ of distinct points of $M$ converging to $P$ (i.e., having $P$ as its sequential limit point) and such that for each point $A$ of $M-P$, there exists an integer $k_a$ such that if $m>k_a$, then $X_m$ separates $A$ and $P$ in $M$.*

(4.1) *Every point of a connected and connected im kleinen point set $M$ is either a cut point, an end point, or a point of some non-degenerate† nodular subset of $M$.*

Let $P$ be any point of $M$ which is neither a cut point of $M$ nor a point of any non-degenerate nodular subset of $M$. I shall proceed to show that $P$ is an end point of $M$. Let $B$ be any point of $M-P$, and let $K$ be the set of all those points of $M$ which separate $P$ and $B$ in $M$. Now $K+B$ is not closed. For if so, then $K+B$ contains at least one limit point $Q$ of the component $N$ of $M-(K+B)$ which contains $P$. Now $Q \neq B$; for if $Q=B$, then $K=0$ and, by (3.8), $P$ and $B$ lie together in a nodular subset of $M$, contrary to supposition. Hence $Q$ belongs to $K$, and therefore $M-Q=M_p(Q)+M_b(Q)$. But then $Q$ and $P$ are not separated in $M$ by any point $X$ of $M$; for if so, then since $X$ would

* For a point $P$ of *any continuum* $M$, the property mentioned in this definition is equivalent to the property of being a point of Menger order one of $M$, i.e., an end point of $M$ in the sense of Menger. This is not true for connected and connected im kleinen sets in general. However, this property is equivalent, in general, to the property of being a point of order 1 of $M$ in the sense that a monotone decreasing sequence $U_1, U_2, U_3, \cdots$ of neighborhoods of $P$ in $M$ exists such that the boundary of $U_i$ in $M$ is a single point $X_i$ and such that

$$P=\prod_1^{\infty}(U_i+X_i).$$

† A set of points is degenerate or non-degenerate according as it does or does not reduce to a single point. This terminology is due to R. L. Moore.

necessarily belong to $N$ and hence to $M_p(Q)$, and $M_b(Q)+Q$ is connected* and contains both $B$ and $Q$ but not $X$, it would follow that $X$ separates $B$ and $P$ in $M$ and hence belongs to $K$. This is absurd, since $X \subset N \subset M-K$. Thus no point of $M$ separates $Q$ and $P$ in $M$. Hence, by (3.8), $P$ and $Q$ lie together in a nodular subset of $M$, contrary to supposition. Therefore it follows that $K+B$ is not closed. Since, by (1.1), $(K+B)+P$ is closed, it follows that $P$ is a limit point of $K$. Accordingly there exists an infinite sequence $P_1, P_2, P_3, \cdots$ of points of $K$ converging to $P$. Now let $A$ be any point whatever of $M-P$. Clearly $(A+P+P_1+P_2+\cdots)-A=S$ is closed. Let $R$ be the component of $M-S$ containing $A$. Then since, by hypothesis, $P$ is not a cut point of $M$, manifestly there exists a point $P_j$ of $S$, distinct from $P$, which is a limit point of $R$. As $P_j$ belongs to $K$, then $M-P_j=M_p(P_j)+M_b(P_j)$. Since $\lim_{n\to\infty} P_n = P$ and $\overline{M_b(P_j)} \cdot P=0$, there exists an integer $k_a$ such that for any integer $m>k_a$, $P_m$ belongs to $M_p(P_j)$. Let $m$ be any integer $>k_a$. Then since $R+P_j+M_b(P_j)$ is connected and contains $A+B$ but not $P_m$, and $P_m$ separates $B$ and $P$ in $M$, it follows that $P_m$ separates $A$ and $P$ in $M$. Thus $P$ is an end point of $M$, and our theorem is proved.

Any non-degenerate maximal nodular subset of a set $M$ will be called a *nodule* of $M$. A subset $E$ of a set $M$ will be called a *nodular element* of $M$ provided that $E$ is either (a) a nodule of $M$, (b) a cut point of $M$, or (c) an end point of $M$. The nodular elements may also be defined as consisting of (1) the cut points of $M$ and (2) the maximal nodular subsets of $M$, the non-degenerate elements of (2) being the nodules of $M$ and the degenerate ones being the end points of $M$.†

(4.2) *If $N_1, N_2, N_3, \cdots$ is any sequence of nodules of a set $M$ having a sequential limiting set $L$, then $L \cdot M$ is either vacuous or it consists of a single point.*

Suppose, on the contrary, that $L$ contains at least two points $A$ and $B$ of $M$. Then since $M$ is connected im kleinen, it readily follows that there exist two integers $i$ and $j$ and two mutually exclusive connected subsets $C_a$ and $C_b$ of $M$ such that $N_i \cdot C_a \neq 0 \neq N_j \cdot C_a$, and $N_i \cdot C_b \neq 0 \neq N_j \cdot C_b$. But then if $H$ denotes the component of $M-N_i$ which contains $N_j - N_i \cdot N_j$ ($N_i \cdot N_j$ contains at most one point), it is clear that $N_i$ contains at least two limit points of $H$, contrary to (3.4).

---

* Cf. Knaster and Kuratowski, *Sur les ensembles connexes*, Fundamenta Mathematicae, vol. 2 (1921), pp. 206–255.

† Compare with R. L. Moore, *Concerning upper semi-continuous collections*, Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 81–88, and also C. Kuratowski, *Quelques applications d'éléments cycliques de M. Whyburn*, Fundamenta Mathematicae, vol. 14 (1929), p. 138.

(4.3) *The nodules of any set $M$ are countable.*

This follows at once from (4.2).

It should be noted that, under the conditions of (4.2), $\lim_{n\to\infty} \delta(N_i)$ may be $\neq 0$. The situation here differs from that existing in the case of the maximal cyclic curves of a continuous curve, due to the fact that the set $M$ may fail to be complete or locally compact in itself. It is also to be noted that, in (4.2), we may allow $N_i$ $(i=1, 2, 3, \cdots)$ to be any connected collection of nodular elements of $M$ (see §6 below), so long as $N_i \cdot N_j$ consists of at most one point, for each $i$ and $j$. This more general result follows by an argument very similar to the proof of (4.2), using (6.4) (below) instead of (3.4).

Proposition (4.1), together with (3.6), gives at once the fact that any connected and connected im kleinen point set may be decomposed into its nodular elements. Thus we may state the following theorem, which follows readily with the aid of (4.3).

(4.4) *If $K$, $H$, and $N$, respectively, denote the sum of all the cut points, end points, and nodules of a set $M$, then* (1) $K+H+N=M$, (2) $K\cdot N$ *is countable, and* (3) $K\cdot H=N\cdot H=0$.

## 5. Simple nodular chains

Let $A_0$ and $B_0$ be any two nodular elements of a connected and connected im kleinen point set $M$, and let $A$ and $B$ respectively denote $A_0$ and $B_0$ or internal points of $A_0$ and $B_0$ according as $A_0$ and $B_0$ are single points or nodules of $M$. Then by the *simple nodular chain* of $M$ from $A_0$ to $B_0$ is meant the subset $X$ of $M$ defined in either of the two following equivalent ways.

(5.1) *If $K$ is the set of all those points of $M$ separating $A$ and $B$ in $M$, then $X=K+A_0+B_0+$all nodules of $M$ each of which contains exactly two points of $K+A+B$.*

(5.2) *$X$ is a subset of $M$ which is* (a) *connected,* (b) *the sum of the elements of some collection of nodular elements of $M$ including $A_0$ and $B_0$, and* (c) *irreducible with respect to properties* (a) *and* (b).

We proceed to show the identity of the sets $X$ defined in (5.1) and (5.2). Denote these sets by $X_1$ and $X_2$ respectively. We first show that $X_1$ is a set satisfying conditions (a), (b), and (c) in (5.2). The set $X_1$ is connected; for if not, it is the sum of two mutually separated sets $U$ and $V$ each of which, it is easily seen, must contain a point of $K+A+B$. Since $K+A+B$ is closed, $U\cdot(K+A+B)=N_1$ and $V\cdot(K+A+B)=N_2$ are closed. One of these sets, say $N_1$, contains $A$. It follows by (1.2) that there exists a first point $Z$ of $N_2$ in the order from $A$ to $B$ in $M$ and likewise a last point $Y$ in $N_1$ which precedes

$Z$ in the order $A$, $B$. But then $Y$ and $Z$ are not separated in $M$ by any point of $M$, since there is no point of $K$ between them. Accordingly, by (3.8), $M$ contains a nodule $N$ containing $Y+Z$. Clearly $N$ contains no other point of $K+A+B$. But then $N \subset X_1$, and this is impossible, since $N$ is connected and $U$ and $V$ are mutually separated. Thus $X_1$ is connected and hence satisfies (a). It satisfies (b) by definition, since $X_1=K+A_0+B_0+$nodules of $M$. Suppose, contrary to (c), that $X_1$ contains a proper subset $Y$ having properties (a) and (b). Clearly $Y \supset K$; but since $Y$ is not $=X_1$, there exists a point $Z$ of $X_1-Y$ which must belong to a nodule $N$ of $M$ which contains two points $U$ and $V$ of $K+A+B$. Now $Y \cdot N \supset U+V$ and since, by (3.5), $Y \cdot N$ is connected, it must contain at least one internal point of $N$. But then, by (3.7), $N \subset Y$, contrary to supposition. Thus $X_1$ has property (c). Hence $X_1$ is a set having properties (a), (b), and (c) in (5.2). That $X_1$ is identical with every set having these properties follows by almost the same argument as just given to show that it has property (c). Thus $X_1=X_2$.

The following two propositions follow readily from the results just established above.

(5.3) *For every two nodular elements $N_1$ and $N_2$ of a set $M$, there exists in $M$ one and only one simple nodular chain $X$ from $N_1$ to $N_2$.*

(5.4) *Every simple nodular chain of a set $M$ is closed in $M$, connected, and connected im kleinen.*

**Note.** For the case where $M$ is closed, i.e., where $M$ is a continuous curve, the nodular elements of $M$ coincide with the cyclic elements of $M$, and the simple nodular chain in $M$ between two nodular elements $N_1$ and $N_2$ of $M$ is identical with the simple cyclic chain in $M$ from $N_1$ to $N_2$ (see my paper *Concerning the structure of a continuous curve*, loc. cit.) and with the *arc-curve $AB$*, where $A$ and $B$ are points of $N_1$ and $N_2$ respectively and are non-cut points of $M$ in case $N_1$ or $N_2$ is non-degenerate, in the sense of W. L. Ayres.* Additional properties of nodular chains similar to those of simple cyclic chains, or arc-curves, could be stated and proved in an analogous way; but to save space the reader is simply referred to §2 of the above mentioned paper of the author's and to the slightly later paper of Ayres.

## 6. Connected collections of nodular elements

A connected subset of a connected and connected im kleinen point set $M$ which is the sum of the elements of some collection of nodular elements of $M$ will be called a *connected collection of nodular elements* of $M$.

* Cf. *Concerning arc-curves and basic sets of a continuous curve*, these Transactions, vol. 30 (1928), pp. 567–578.

(6.1) *If $N_1$ and $N_2$ are nodular elements of a set $M$ belonging to a connected collection $H$ of nodular elements of $M$, then $H$ contains the simple nodular chain $X$ of $M$ between $N_1$ and $N_2$.*

Let $A$ and $B$ be points of $N_1$ and $N_2$ respectively, and let them be internal points if $N_1$ and $N_2$ or either is a nodule of $M$. Let $K$ be the set of all points of $M$ each of which separates $A$ and $B$ in $M$. Now suppose that $H$ does not contain $X$. Then $X-H\cdot X$ contains a point $P$. Since $H$ is connected and contains $A+B$, therefore it contains $K$. Hence, by (5.1), $P$ belongs to a nodule $N$ of $M$ containing two points $U$ and $V$ of $K+A+B$. But since, by (3.5), $N\cdot H$ is connected, then $H$ contains an internal point of $N$; and hence, by (3.7), $H \supset N \supset P$, contrary to supposition. This contradiction proves our theorem.

(6.2) *If $H$ is any connected collection of nodular elements of a set $M$, then every point of $\overline{H}\cdot M-H$ is a nodular element of $M$ (and hence is either a cut point or an end point of $M$).*

We proceed to show that each point of $\overline{H}\cdot M-H$ is either a cut point or an end point of $M$. If this is not so, then some point $P$ of $\overline{H}\cdot M-H$ is an internal point of some nodule $N$ of $M$. Now $H\cdot N\neq 0$; for if not, it would follow by (3.4) that one and only one point $Q$ of $N$, different from $P$, is a limit point of $H$. Hence $H$ contains a point $Q$ of $N$; and since $P$ is a limit point of $H$, it follows readily that $N\cdot H$ contains at least two points. And since, by (3.5), this set $N\cdot H$ is connected, therefore it contains an internal point of $N$. But then, by (3.7), $H \supset N \supset P$, contrary to the fact that $P \subset \overline{H}\cdot M-H$. Thus $P$ is either a cut point or an end point of $M$, and our theorem is established.

(6.3) *Under the hypothesis of* (6.2), *$\overline{H}\cdot M$ is a connected collection of nodular elements of $M$.*

(6.4) *If $H$ is any connected collection of nodular elements of a set $M$, and $N$ is any component of $M-H$, then $M\cdot\overline{H}\cdot\overline{N}$ consists of exactly one point.*

That this set contains at least one point follows from the fact that $M$ is connected im kleinen. Suppose, contrary to (6.4), that it contains two points $A$ and $B$. Since $H+A+B$ and $N+A+B$ are connected subsets of $M$, and $N\cdot H=0$, clearly no point of $M$ separates $A$ and $B$ in $M$. Consequently, by (3.8), $A+B$ lies in a single nodule $C$ of $M$. By (3.5), $C\cdot(H+A+B)$ is connected, and since it contains $A+B$, it follows by (3.7) that $H \supset C$. But then $(N+A+B)\cdot C=A+B$, a disconnected set, contrary to (3.5). Thus our theorem is established.

(6.5) *If $H$ is any connected collection of nodular elements of a set $M$ and $W$ is any connected subset of $M$, then $H \cdot W$ is either vacuous or connected.*

Let $A$ and $B$ be any two points of $H \cdot W$, and let $A_0$ and $B_0$ respectively denote nodular elements of $M$ belonging to $H$ and containing $A$ and $B$. By (6.1), the simple nodular chain $X$ in $M$ from $A_0$ to $B_0$ is a subset of $H$. Now since $X$ is closed in $M$ and is a connected collection of elements of $M$, it follows by (6.4) and (2.1) that $W \cdot X$ is connected. And since $W \cdot X$ contains $A+B$ and is a subset of $H \cdot W$, and $A$ and $B$ are any two points of $H \cdot W$, it follows that $H \cdot W$ is connected.

(6.6) *Every connected collection $H$ of nodular elements of a set $M$ is connected im kleinen.*

Since $M$ is connected im kleinen, then for each point $P$ of $M$ and each $\epsilon > 0$, a $\delta_\epsilon > 0$ exists such that for each point $Q$ of $H$ with $\rho(Q, P) < \delta_\epsilon$, a connected subset $C$ of $M$ exists with $P+Q \subset C$ and $\delta(C) < \epsilon$. And since, for every such set $C$, by (6.5), $C \cdot H$ is connected, it follows that $H$ is connected im kleinen.

## 7. Some applications; denodular sets

In this section the results of the preceding sections will be applied to yield some interesting (new) results concerning the structure of certain special types of connected and connected im kleinen sets.

A set $M$ which satisfies any one of the following equivalent conditions will be called a *denodular set.*

(7.1) *$M$ contains no nodules, i.e., all its nodular elements are points.*

(7.2) *Every two points of $M$ are separated in $M$ by some third point of $M$.*

(7.3) *Every point of $M$ is either a cut point or an end point.*

The equivalence of these three properties is easily established with the aid of the results of the preceding sections.

(7.4) *Every connected set $H$ of degenerate nodular elements of a set $M$ (i.e., every connected set of cut points and end points of $M$) is a denodular set.*

By (6.6), $H$ is connected im kleinen and hence is itself a set $M$. That $H$ is denodular follows immediately now from (7.1).

(7.5) *Every denodular set $M$ is arcwise connected; indeed it is uniquely arcwise connected.*

Let $A$ and $B$ be any two points of $M$. By (5.2), $M$ contains a simple

nodular chain $X$ from $A$ to $B$. By (5.1) it follows that $X \equiv K+A+B$, where $K$ is the set of all points of $M$ separating $A$ and $B$ in $M$. And since by (1.1), $K+A+B$, or $X$, is compact and closed, it is then a simple continuous arc from $A$ to $B$. That $X$ is the only arc in $M$ from $A$ to $B$ follows from (5.3).

(7.6) *Every connected set of cut points and end points of any connected and connected im kleinen set $M$ is uniquely arcwise connected.*

This is an immediate consequence of (7.4) and (7.5).

(7.7) LEMMA. *Let $A$ and $B$ be any two points of a connected and arcwise connected im kleinen set $M$, let $C$ be the collection of all connected subsets of $M$ containing both $A$ and $B$, let $T$ be the collection of all arcs in $M$ from $A$ to $B$, and let $L$ be the greatest lower bound of the number aggregate $[\delta(c)]$, where $c$ is an element of $C$. Then $L$ is also the greatest lower bound for the aggregate $[\delta(t)]$, where $t$ is an element of $T$.*

Let $\epsilon$ be any positive number. By the hypothesis, a connected subset $c$ of $M$ exists with $\delta(c) < L+\epsilon/4$. Since $M$ is connected im kleinen, then for each point $X$ of $c$, a connected open subset $H_x$ of $M$ exists containing $x$ and with $\delta(H_x) < \epsilon/4$. Then if $H = \sum H_x$, it is clear that $H$ is a connected open subset of $M$, and $\delta(H) < L+\epsilon$. But* $H$ is arcwise connected; and since it contains $A+B$, it therefore contains an arc $t$ in $M$ from $A$ to $B$. Then $t$ is an element of $T$ and $\delta(t) < L+\epsilon$; accordingly, $L$ is the greatest lower bound for $[\delta(t)]$, and our lemma is proved.

(7.8) *If every nodule of a set $M$ is arcwise connected im kleinen, then $M$ itself is arcwise connected im kleinen. Hence $M$ is also arcwise connected.*

Let $A$ be any point of $M$. For each $\epsilon > 0$, a $\delta_{\epsilon a} > 0$ exists such that for each point $x$ of $M$ with $\rho(A, x) < \delta_{\epsilon a}$ a connected subset $C$ of $M$ exists with $A+x \subset C$ and $\delta(C) < \epsilon/8$. Let $B$ be any such point $x$ of $M$. Let $K$ be the set of all points of $M$ separating $A$ and $B$ in $M$, and let $X$ be the simple nodular chain in $M$ from $A$ to $B$. Let the nodules of $M$ in $X$ be ordered (they are countable by (4.3)): $N_1, N_2, N_3, \cdots$. For each $i > 0$, let $A_i$ and $B_i$ be the two points (see (5.1)) of $N_i \cdot (K+A+B)$, and let $L_i$ be the greatest lower bound of $[\delta(c_i)]$ where $C_i$ is the collection of all connected subsets of $N_i$ containing $A_i+B_i$, and $c_i$ is an element of $C_i$. For each $i$, by (7.7), $N_i$ contains an arc $t_i$ from $A_i$ to $B_i$ with

---

* See G. T. Whyburn, *Concerning the complementary domains of continua*, Annals of Mathematics, vol. 29, pp. 399–411, Theorem 12. This theorem, together with its proof, holds equally well in the space here considered.

(i) $$\delta(t_i) < L_i + \epsilon/(8i).$$

Let

(ii) $$W = K + A + B + \sum_{i=1}^{\infty} t_i.$$

Then $W$ is an arc from $A$ to $B$, as will now be demonstrated. Clearly $W$ is connected and irreducibly connected between $A$ and $B$. It is also closed; for if not, then since $K+A+B$ is closed, there would exist a positive number $d$ and an infinite sequence $t_{n_1}, t_{n_2}, \cdots$ of the arcs $[t_i]$ with

(iii) $$\delta(t_{n_i}) > d,$$

for every $i$, and furthermore such that there exists a sequential limiting set $L$ of this sequence containing a point $P$ of $K+A+B$. Since $M$ is connected im kleinen at $P$, a $\delta>0$ exists such that for each point $x$ of $M$ with $\rho(P, x)<\delta$, a connected subset $H$ of $M$ exists containing $P+x$ and with $\delta(H)<d/8$. Since, as is easily established, $\lim_{i\to\infty} \rho(A_i, B_i)=0$, and since by (4.2), $L\cdot M = P$, it readily follows that an integer $n_j$ exists such that

(iv) $$\epsilon/(8n_j) < d/8$$

and such that $\rho(A_{n_j}+B_{n_j}, P)<\delta$. There exists, then, a connected subset $H$ of $M$ with $(A_{n_j}+B_{n_j}+P) \subset H$ and $\delta(H)<d/4$. Now by (3.5), $H\cdot N_{n_j}$ is connected, and since $\delta(H\cdot N_j)<d/4$ and $A_{n_j}+B_{n_j} \subset H\cdot N_j$, it follows that $L_{n_j} \leqq d/4$. Hence by (i) and (iv), it follows that

(v) $$\delta(t_{n_j}) < L_{n_j} + \epsilon/(8n_j) \leqq d/4 + d/8 < d.$$

But (iii) and (v) are contradictory. Therefore $W$ is closed and hence is an arc from $A$ to $B$. It remains to show that $\delta(W)<\epsilon$. Now since a connected subset $C$ of $M$ exists with $A+B \subset C$ and $\delta(C)<\epsilon/8$, and since by (3.5), $C\cdot N_i$ is connected and $C\cdot N_i \supset A_i+B_i$ for every $i$, it follows that $L_i<\epsilon/8$ for every $i$. Hence by (i), $\delta(t_i)<\epsilon/8+\epsilon/(8i)\leqq\epsilon/4$ for every $i$. Now $\delta(K+A+B)<\epsilon/8$, for $K+A+B \subset C$; and since every arc $t_i$ has two points in $K+A+B$, it follows from (ii) that

$$\delta(W)<2\cdot\epsilon/4\epsilon/8<\epsilon.$$

Thus $M$ is arcwise connected im kleinen at every point. That $M$ is also arcwise connected follows now by the theorem of the author's quoted above.

**Definition.** A connected and connected im kleinen point set $M$ will be said to be locally denodular if for each point $P$ of $M$ a neighborhood $R$ of $P$ exists such that the component of $M\cdot R$ containing $P$ is a denodular set. This

notion is analogous to Menger's notion of a Baum im kleinen.*

(7.9) *If every nodule of a set $M$ is locally denodular, then $M$ is arcwise connected and arcwise connected im kleinen.*

It follows by (7.5) that every nodule of $M$ is arcwise connected im kleinen. Hence by (7.8), $M$ is arcwise connected im kleinen and arcwise connected.

**Definition.** A collection $G$ of subsets of a set $M$ will be called an $E$-collection provided that if $g_1, g_2, g_3, \cdots$ is any sequence of elements of $G$ having a non-vacuous sequential limiting set $L$ with $L \cdot M \neq 0$, then $\lim_{n\to\infty} \delta(g_n) = 0$.

As pointed out above, the collection $G$ of all the nodules of a set $M$ may fail to be an $E$-collection, although it is true that the collection of all maximal cyclic curves of a continuous curve is an $E$-collection. As might be expected, it is true that if we stipulate that the collection $G$ of nodules is an $E$-collection, we get a set $M$ which is more completely analogous, from the standpoint of its nodular elements, to a continuous curve; and some known theorems about continuous curves extend to such sets $M$ which would not extend to sets $M$ in general. As an example, we state the following one.

(7.10) *If the collection of all nodules of a set $M$ is an $E$-collection, then in order that $M$ should be arcwise connected it is necessary and sufficient that each nodule of $M$ be arcwise connected.*

This theorem is proved by a rather obvious modification of the proof of (7.8). A number of other theorems of the same type could also be stated for such sets $M$. The reader is referred to §6 of my paper *Concerning the structure of a continuous curve*. It is to be noted that if $M$ is any connected subset of a Menger regular curve, then it is connected im kleinen and the collection of all nodules of $M$ has property $E$.

## 8. Linear arrays of nodular elements

**Definition.** The connected collection $H$ of nodular elements of a set $M$ will be called a *linear array* of elements of $M$ provided every nodular element of the set $H$ itself which cuts $H$ cuts it into exactly two components. It is readily seen that not more than two nodular elements of a linear array $H$ of elements can fail to cut $H$. If every element of the linear array $H$ cuts $H$, then $H$ is called an *open chain* of nodular elements of $M$; if $H$ itself has more than one nodular element, and if every element, save one element $E$, of $H$ cuts $H$, then $H$ will be called a *radial chain* of nodular elements of $M$ and will

* Cf. *Über reguläre Baumkurven*, Mathematische Annalen, vol. 96 (1926), pp. 572–582.

be said to emerge from $E$; if every element, save two elements $A_0$ and $B_0$, of $H$ cuts $H$, then $H$ is identically the simple nodular chain in $M$ from $A_0$ to $B_0$. (See §5.)

If $E$ is any nodular element of an open chain $L$ of nodular elements of a set $M$, then $L$ is the sum of two radial chains $r_1$ and $r_2$ emerging from $E$ and having only $E$ in common; furthermore, there exist two sequences $A_1, A_2, A_3, \cdots$ and $B_1, B_2, B_3, \cdots$ of cut points of $L$ belonging to $r_1$ and $r_2$ respectively and such that neither $[A_i]$ nor $[B_i]$ has a limit point in $L$ and if, for each $i$, $X_i$ is the simple nodular chain in $M$ from $A_i$ to $B_i$, then $L=\sum X_i$ and, for each $i$, $X_i \subset X_{i+1}$. If $L$ is any open chain in a set $M$, then $M\cdot\overline{L}-L$ contains at most two points; and $M\cdot\overline{L}$ is either a simple nodular chain, a radial chain, or an open chain in $M$ according as $M\cdot\overline{L}-L$ contains two points, one point, or no points. If $L$ is any open chain of nodular elements of a set $M$, then there exists in $M$ a maximal open chain of nodular elements containing $L$, i.e., an open chain of elements which is saturated with respect to the property of being an open chain of nodular elements of $M$; if $r$ is any radial chain of nodular elements of $M$, then there exists in $M$ either a maximal radial chain or a maximal open chain (sometimes both) containing $r$; if $X$ is any simple nodular chain in $M$, then there exists in $M$ either a maximal simple nodular chain, a maximal radial chain or a maximal open chain containing $X$. If $U$ is any maximal simple nodular chain of a set $M$, or is any maximal radial chain or maximal open chain of $M$ which is closed in $M$, then $U$ is a *maximal linear array* in $M$, i.e., no linear array in $M$ contains $U$ as a proper subset. Clearly any maximal linear array in $M$ must be closed in $M$, and indeed, if $H$ is any linear array of nodular elements in a set $M$, then there exists in $M$ a maximal linear array which contains $H$. The facts just enumerated can be established quite readily by methods of argument which are well known in point set theory.

If $M$ is a denodular set, it is easily seen that every open chain of nodular elements of $M$ is an ordinary open curve, or what is equivalent, an arc segment, every radial chain in $M$ is an ordinary ray, and every simple nodular chain is an ordinary simple continuous arc. Thus if $M$ is denodular, and $L$ is any open curve in $M$, then $M\cdot\overline{L}-L$ is either vacuous, one point, or two points. And if $L$ is any maximal open curve in $M$, then each point, if any, of $M\cdot\overline{L}-L$ is an end point of $M$; and $\overline{L}\cdot M$ is either $\equiv L$, or is a maximal ray in $M$, or is a maximal arc in $M$, according as $M\cdot\overline{L}-L$ is vacuous, one point, or two points. Furthermore, if $L$ is any open curve in $M$, there exists in $M$ a maximal open curve containing $L$; if $r$ is any ray in $M$, there exists in $M$ either a maximal ray or a maximal open curve containing $r$; and if $AB$ is any arc in $M$, there exists in $M$ either a maximal arc, a maximal ray, or a maximal open curve containing $AB$.

(8.1) *Let $M$ be any connected and connected im kleinen set. Then there exists in $M$ a countable sequence of point sets $S_1, S_2, S_3, \cdots$ such that* (1) *$S_1$ is a maximal linear array of nodular elements of $M$,* (2) *for each $n>1$, $S_n$ has exactly one point $A_n$ in common with $\sum_{i=1}^{n-1} S_i$, $S_n - A_n$ is a subset of some component $N_n$ of $M - \sum_{i=1}^{n-1} S_i$ which has $A_n$ as its limit point in $\sum_{i=1}^{n-1} S_i$, and $S_n$ is a maximal linear array of nodular elements in the set $N_n + A_n$ [hence either $S_n$ is a radial chain emerging from $A_n$ (and lying otherwise in $N_n$) or else there exists a nodular element $B_n$ such that $S_n$ is the simple nodular chain in $M$ from $A_n$ to $B_n$ (a single nodule may be a simple nodular chain)], and* (3) *every point, if any, of $M - \sum_1^\infty S_n$ is an end point of $M$.*

Using the fact that $M$ is separable, this proposition may be established on the basis of the preceding results by an argument which is by no means difficult.

(8.2) *Let $G$ be the sum of all cut points and all the end points of any connected and connected im kleinen point set $M$. Then there exists a subset $H$ of the universal acyclic continuous curve* $W$ of Wazewski and a biunivalued transformation $T$ such that* (1) *$T(G)=H$,* (2) *$T$ is continuous, and $T^{-1}$ is continuous over every arc in $H$.*

This proposition may be proved on the basis of the properties of the nodular elements of $M$ by an argument not essentially different from that used in my paper *On the set of all cut points of a continuous curve*† to prove an analogous though stronger theorem concerning the set of all cut points and end points of a continuous curve. It should be noted that it is not always possible for the transformation $T$ to be bicontinuous, even in case $M$ is a denodular set. For example, let $A$ denote the $X$-axis, and for each integer $n \neq 0$, let $L_n$ denote the line $x = 1/n$; and finally, let $M = A + \sum L_n$. Then clearly the set $G$, which is $\equiv M$, is not homeomorphic with any subset of an acyclic continuous curve. It seems to be true, however, that for sets $M$ having the property that if $X_1, X_2, \cdots$ is any sequence of mutually exclusive simple nodular chains in $M$ having a sequential limiting set which contains a point of $M$, then $\lim_{n\to\infty} \delta(X_n) = 0$, it is true that the set $G$ is homeomorphic with some subset $H$ of an acyclic continuous curve.

If $M$ is denodular, then $G \equiv M$ and proposition (8.2) gives the following interesting corollary:

* This is a compact plane acyclic continuous curve $W$ such that every acyclic continuous curve is homeomorphic with some subset of $W$. See T. Wazewski, *Sur les courbes de Jordan non renfermant aucune courbe simple fermée de Jordan*, Annales de la Société Polonaise Mathématique, vol. 2 (1923), p. 169.

† Fundamenta Mathematicae, vol. 15, pp. 185–194.

(8.2a) *For every denodular set $D$ there exists a subset $H$ of an acyclic continuous curve and a biunivalued transformation $T$ such that* (1) $T(D)=H$, *and* (2) $T$ *is continuous, and* $T^{-1}$ *is continuous over every arc in* $H$.

## 9. Conclusion; some unsolved problems

The author is quite well aware that the above group of propositions is by no means exhaustive on this subject. An attempt has been made, in general, only to include those theorems and proofs which are essential in the construction of the foundation for the central idea of the paper. One who has followed through the results in the above sections and who is familiar with the cyclic elements of a continuous curve cannot fail, I believe, to see the pattern, beautiful in and because of its simplicity, of the connected and connected im kleinen point set neatly arranged into nodular elements with respect to which the structure of the whole set is so like that of an acyclic continuous curve. Once this pattern is grasped, it is a simple matter to observe and establish many other properties of and relations among the nodular elements. In the course of this investigation, however, there have occurred to me two problems of an entirely different nature, which I have been unable to solve and whose solution would, I believe, be of considerable interest and importance. They are as follows.

(9.1) **Problem.** *Do the nodules of a set $M$ possess any property which might be called "cyclic"?* It is known, of course, that if $M$ is a continuous curve, then every nodule $C$ of $M$, or what is the same thing, every maximal cyclic curve $C$ of $M$ is *cyclicly connected*, i.e., every two points of $C$ lie together on a simple closed curve in $C$. Do we have any sort of analogue to this property in a set $M$ in general? Or, more specifically, if a set $M$ has no cut point, then is it true that for every two points $A$ and $B$ of $M$, two connected subsets $H$ and $K$ of $M$ exist with $H\cdot K\equiv A+B$? If so, can $H$ and $K$ be chosen so that $H-(A+B)$ and $K-(A+B)$ are mutually separated? The terms nodule and nodular element were introduced and used in this paper because of the lack of knowledge of the existence of any such property to justify the use of the terms *cyclic subset* and *cyclic element*, as was done for a continuous curve.

(9.2) **Problem.** *In order that the bounded set $M$ lying in a plane $Z$ should cut $Z$ in the weak sense, is it necessary* (it is sufficient) *that some nodule of $M$ cut $Z$ in the weak sense?* Or less generally, is it true that no bounded denodular set $M$ lying in a plane $Z$ can cut $Z$ in the weak sense? It also seems probable to me that if the set $M$ lying in a plane $Z$ does not cut $Z$ in the weak sense, then either $M$ is denodular or else, for each nodule $N$ of $M$, there exists a domain $R$ such that $R\subset N\subset\overline{R}$.

University of Texas,
Austin, Texas

## ERRATA, VOLUME 26

J. H. M. WEDDERBURN, *Algebras which do not possess a finite basis.*
Page 420, lines 1 and 3: for "a modulus" read "no modulus."

## ERRATA, VOLUME 31

J. F. RITT, *Algebraic combinations of exponentials.*
Page 665, line 15, for "$\psi(z)e^{-\lambda_1 z}\phi(z)$" read "$\psi(z) = e^{-\lambda_1 z}\phi(z)$."
Page 676, line 16, for "$S$" read "a subsector of $S$."
Page 679, line 2, for "$z$" read "$z/(2\pi)$."